

# Threshold estimation for stochastic processes with small noise

Yasutaka Shimizu\*

*Department of Applied Mathematics, Waseda University  
Japan Science and Technology Agency, CREST*

February 25, 2015

## Abstract

Consider a process satisfying a stochastic differential equation with unknown drift parameter, and suppose that discrete observations are given. It is known that a simple least squares estimator (LSE) can be consistent, but unstable under finite samples when the noise process has jumps. We propose a filter to cut large shocks from data, and construct the same LSE from data selected by the filter. The proposed estimator can be asymptotically equivalent to the usual LSE, whose asymptotic distribution strongly depends on the noise process. However it is interesting to note that it could be asymptotically normal by choosing the filter suitably if the noise is a Lévy process. We try to justify this phenomenon theoretically.

*Key words:* stochastic differential equation, semimartingale noise, small noise asymptotics, drift estimation, threshold estimator, mighty convergence.

*MSC2010:* **62F12**, **62M05**; 60G52, 60J75

## 1 Introduction

Let  $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})$  be a stochastic basis, on which an  $\mathbb{R}^d$ -valued stochastic process  $X$  is defined via the stochastic integral equation

$$X_t^\epsilon = x + \int_0^t b(X_s^\epsilon, \theta_0) ds + \epsilon \cdot Q_t^\epsilon, \quad (1.1)$$

where  $x \in \mathbb{R}^d$ ,  $\epsilon > 0$ , and  $\theta_0$  is an unknown parameter that belongs to a parameter space  $\Theta_0$ , which is an open bounded, convex subset of  $\mathbb{R}^p$ ; we put  $\Theta := \overline{\Theta_0}$ , the closure of  $\Theta_0$ ;  $b$  is a measurable function on  $\mathbb{R}^d \times \Theta$ ;  $Q^\epsilon$  is a stochastic process such that

$$Q_0^\epsilon = 0 \quad a.s., \quad \sup_{t \in [0,1]} |Q_t^\epsilon - Q_t| \rightarrow 0 \quad a.s. \quad (1.2)$$

---

\*E-mail: shimizu@waseda.jp

where  $Q$  is a semimartingale with the Doob-Meyer decomposition

$$Q_t = A_t + M_t \quad (t \geq 0), \quad A_0 = M_0 = 0 \quad a.s., \quad (1.3)$$

where  $A$  is a process with finite variation, and  $M$  is an  $\mathcal{F}_t$ -local martingale. Some typical examples for  $Q^\epsilon$  is given in Section 2.2. We suppose that the process  $X^\epsilon = (X_t^\epsilon)_{t \in [0,1]}$  is observed discretely in time:  $\{X_{t_k^n}\}_{k=0}^n$  with  $t_k^n = k/n$  (the index  $\epsilon$  is abbreviated), from  $[0,1]$ -interval. We denote by  $\Delta_k^n X := X_{t_k^n} - X_{t_{k-1}^n}$  and  $\Delta_n := t_k^n - t_{k-1}^n = 1/n$ . Our interest is to estimate the value of the parameter  $\theta_0$  from the discrete samples under that  $n \rightarrow \infty$  as well as  $\epsilon \rightarrow 0$ : *small noise asymptotics*.

There are some practical advantages in small noise asymptotics:

- *Statistical point of view*: the drift estimation by samples from a fixed finite time interval is justifiable under relatively mild conditions. Since we need to observe the process ‘long’ time to estimate the drift, we usually assume the asymptotics that the terminal time of observations goes to infinity, under which some technical conditions such as “ergodicity” or “uniform moment conditions” for the process need to be assumed. Small noise asymptotics can avoid those conditions that is sometimes difficult to check in practice; see also remarks in Section 2.2.
- *Computational point of view*: approximation of functionals of the process as  $\epsilon \rightarrow 0$  is often available in relatively easy-to-calculate form as in, e.g., Yoshida [36] and Pavlyukevich [23] among others, which is well applied to finance and insurance; see Takahashi [31], Kunitomo and Takahashi [12], Takahashi and Yoshida [32], Uchida and Yoshida [35], Pavlyukevich [22] and references therein.

That is why, using the small noise model is convenient to deal with both applications and statistical inference at the same time.

Sampling problems for stochastic differential equations with small noise has been well studied by many authors in both theoretical and applied point of views. Some earlier works for *small-diffusion model* are found in the papers by Kutoyants [13, 14], Genon-Catalot [6] and Laredo [15], and they have been developed in some directions by several authors: e.g., martingale estimating functions are studied by Sørensen [29], efficient estimation is investigated by Sørensen and Uchida [30], Gloter and Sørensen [8]; see also Uchida [33, 34], and asymptotic expansion approach is initiated by Yoshida [36, 37], see also Uchida and Yoshida [35], among others. Although those works are due to diffusion noise, more general noise model are also considered recently. For example, Long [16] and Ma [18] investigate the drift estimation of a Lévy driven Ornstein-Uhlenbeck process; see also Long [17], and Long et al. [20] deal with the inference for non-linear drift under the small semimartingale noise. In our paper, we do not require that the noise is a semimartingale, but ‘approximately’ a semimartingale in the sense of (1.2).

The goal of this paper is statistical inference for the drift of the process under the asymptotics that noise vanishes, which is the same motivation as in Long et al. [20], but we get further into more accurate estimation than those in the sense of finite sample performance. Long et al. [20] consider the case where  $Q^\epsilon \equiv L$  is a Lévy process (although it

can be extended to a case of a semimartingale) independent of the dispersion parameter  $\epsilon$ , and investigate the asymptotic behavior of the *least squares-type estimator (LSE)* defined by

$$\hat{\theta}_{n,\epsilon}^{LSE} := \arg \min_{\theta \in \Theta} \Psi_{n,\epsilon}(\theta), \quad (1.4)$$

where

$$\Psi_{n,\epsilon}(\theta) = \epsilon^{-2} \Delta_n^{-1} \sum_{k=1}^n |\Delta_k^n X - b(X_{t_{k-1}^n}, \theta) \cdot \Delta_n|^2 \quad (1.5)$$

They show that the minimum contrast estimator  $\hat{\theta}_{n,\epsilon}^{LSE}$  is  $\epsilon^{-1}$ -consistent with the limit of a Lévy functional as  $\epsilon \rightarrow 0$  and  $n \rightarrow \infty$  with  $n\epsilon \rightarrow \infty$  under some mild conditions on the function  $b$ ; see Theorems 4.1 and 4.2 in [20]. The asymptotic distribution generally has a fat-tail, that leads us unsatisfactory performance even if  $\epsilon$  is small enough; see numerical results in [20], or Section 3 below. This would be due to ‘large’ shocks by the driving noise. It will be easy to imagine that a ‘large’ jump of  $Q^\epsilon$  makes much impact to the direction of drift, and make the drift estimation unstable. Therefore, cutting such ‘large’ jumps could improve the performance. That is why, we consider the *threshold-type* estimator defined as follows:

$$\hat{\theta}_{n,\epsilon} := \arg \min_{\theta \in \Theta} \Phi_{n,\epsilon}(\theta), \quad (1.6)$$

where

$$\Phi_{n,\epsilon}(\theta) = \epsilon^{-2} \Delta_n^{-1} \sum_{k=1}^n |\Delta_k^n X - b(X_{t_{k-1}^n}, \theta) \cdot \Delta_n|^2 \mathbf{1}_{\{|\Delta_k^n X| \leq \delta_{n,\epsilon}\}}, \quad (1.7)$$

$\delta_{n,\epsilon}$  is a positive number, which is a threshold to eliminate ‘large’ shocks causing bias to drift estimation. That is, the indicator  $\mathbf{1}_{\{|\Delta_k^n X| \leq \delta_{n,\epsilon}\}}$  plays a role of a filter to split increments with ‘large’ and ‘small’ magnitude of shocks; see Shimizu [26], or Shimizu and Yoshida [28] for the fundamental idea of those filters. It would be intuitively clear that if  $\delta_{n,\epsilon} \rightarrow \infty$  then  $\hat{\theta}_{n,\epsilon}^{LSE}$  and  $\hat{\theta}_{n,\epsilon}$  can be asymptotically equivalent. However, it does make sense in practice that  $\delta_{n,\epsilon}$  is relatively small number such that  $\delta_{n,\epsilon} \rightarrow 0$  since it plays a role of a filter to cut ‘large’ shocks. Below, we will show that  $\hat{\theta}_{n,\epsilon}$  can be asymptotically equivalent to  $\hat{\theta}_{n,\epsilon}^{LSE}$  even if  $\delta_{n,\epsilon} \rightarrow 0$  by choosing the sequence  $\delta_{n,\epsilon}$  carefully, and it has better finite-sample performance than that of  $\hat{\theta}_{n,\epsilon}^{LSE}$  as well. Furthermore, we will show the *mighty convergence* (the convergence of moments) for  $\hat{u}_{n,\epsilon} := \epsilon^{-1}(\hat{\theta}_{n,\epsilon} - \theta_0)$ , which is a stronger result than those in [20]:

$$\mathbb{E}[f(\hat{u}_{n,\epsilon})] \rightarrow \int_{\mathbb{R}^p} f(z) \mathcal{L}_u(dz),$$

as  $\epsilon \rightarrow 0$  and  $n \rightarrow \infty$  for every continuous function  $f$  of at most polynomial growth, where  $\mathcal{L}_u$  is the asymptotic distribution of  $\hat{u}_{n,\epsilon}$ .

As is described above, the asymptotic distribution  $\mathcal{L}_u$  is generally not normal unless the limiting process  $Q$  is a Wiener process. However, it is interesting to note that we sometimes encounter a phenomenon that  $\hat{u}_{n,\epsilon}$  seems asymptotically normal in numerical studies. This may indicate that our filtered LSE could also be asymptotically normal if we choose  $\delta_{n,\epsilon}$  in a different way as previous, otherwise we may just observe it as an ‘approximate’ phenomenon when some ‘good’ conditions are eventually satisfied. We will try to explain these phenomena theoretically according to some cases where jump activity is finite or infinite.

The paper is organized as follows. In Section 2, we prepare notation and assumptions, and present the main results under discrete samples. In particular, Section 2.4 is devoted to investigating some technical conditions for  $Q$ . We will give some easy-to-check sufficient conditions for those when the noise  $Q$  is a Lévy process. In Section 3, we will show an advantage of our estimator compared with the usual LSE via numerical studies, and we further observe the asymptotic distribution seems normal. Finally, we will investigate those asymptotic phenomena theoretically in Section 4. The proofs of main theorems are given in Section 5.

## 2 Main results

### 2.1 Notation and assumptions

We use the following notation:

- For a process  $Y = (Y_t)_{t \in [0,1]}$ ,  $\|Y\|_{L^p} = (\mathbb{E}|Y_1|^p)^{1/p}$  ( $p > 0$ ) and  $\|Y\|_* := \sup_{t \in [0,1]} |Y_t|$ .
- $\partial_{z_1 \dots z_m} := (\partial/\partial z_1) \dots (\partial/\partial z_m)$ . Moreover,  $\partial_z^\alpha := \partial_{z_1}^{\alpha_1} \dots \partial_{z_m}^{\alpha_m}$  for  $z = (z_1, \dots, z_m)^\top$ , and a multi-index  $\alpha = (\alpha_1, \dots, \alpha_m)$ , where  $\top$  is the transpose.
- Given a multilinear form  $\mathcal{M} = \{M^{(i_1, \dots, i_K)} : i_k = 1, \dots, d_k; k = 1, \dots, K\} \in \mathbb{R}^{d_1} \otimes \dots \otimes \mathbb{R}^{d_K}$  and vector  $u_k = (u_k^{(i)})_{i \leq d_k} \in \mathbb{R}^{d_k}$ , we write

$$\mathcal{M}[u_1, \dots, u_K] = \sum_{i_1=1}^{d_1} \dots \sum_{i_K=1}^{d_K} M^{(i_1, \dots, i_K)} u_1^{(i_1)} \dots u_K^{(i_K)}.$$

Note that the above form is well-defined when the  $j$ th dimension of  $\mathcal{M}$  (the number of  $i_j$ ) and that of  $u_j$  are the same. When some of  $u_k$  is missing in “ $\mathcal{M}[u_1, \dots, u_K]$ ”, the resulting form is regarded as a multilinear form again; e.g.,  $\mathcal{M}[u_3, \dots, u_K] \in \mathbb{R}^{d_1} \otimes \mathbb{R}^{d_2}$ . For example, when  $\mathcal{M}$  is a vector  $M = (M_i)_{1 \leq i \leq d}$ ,  $\mathcal{M}[x] = M^\top x$  for  $x \in \mathbb{R}^d$ , which is the inner product, and when  $\widetilde{\mathcal{M}}$  is a matrix  $M = (M_{ij})_{1 \leq i \leq d_1; 1 \leq j \leq d_2}$ ,  $\widetilde{\mathcal{M}}[u, v]$  for  $u = (u_i)_{1 \leq i \leq d_1}$  and  $v = (v_i)_{1 \leq i \leq d_2}$  is the quadratic form  $u^\top M v$ , and  $\widetilde{\mathcal{M}}[v] = M v \in \mathbb{R}^{d_1}$  or  $\widetilde{\mathcal{M}}[u] = M^\top u \in \mathbb{R}^{d_2}$ , among others. The correspondences of dimensions will be clear from the context. We also use the notation  $\mathcal{M}[u^{\otimes K}] := \mathcal{M}[u_1, \dots, u_K]$  when  $u_1 = \dots = u_K$ .

- For  $a = (a_1, \dots, a_m)^\top \in \mathbb{R}^m$ ,  $\nabla_a = (\partial_{a_1}, \dots, \partial_{a_m})$ . Moreover, we denote by  $\nabla_a^k := \nabla_a \otimes \nabla_a^{k-1}$  with  $\nabla_a^0 \equiv 1$ . That is,  $\nabla_a^k$  forms a multilinear form, e.g.,  $\nabla_a^2 = \nabla_a^\top \nabla_a$  in a matrix form.
- For a multilinear form  $\mathcal{M}$ ,  $|\mathcal{M}|^2$  denotes the sum of the squares of each element of  $\mathcal{M}$ .
- $C$  is often used as an universal positive constant that may differ from line to line. Moreover, we write  $a \lesssim b$  if  $a \leq Cb$  almost surely.
- $C^{k,l}(\mathbb{R}^d \times \Theta; \mathbb{R}^q)$  denotes the space of functions  $f(x, \theta) : \mathbb{R}^d \times \Theta \rightarrow \mathbb{R}^q$  that is  $k$  and  $l$  times differentiable with respect to  $x$  and  $\theta$ , respectively. Moreover,  $C_\uparrow^{k,l}(\mathbb{R}^d \times \Theta; \mathbb{R}^q)$  denotes a subclass of  $f \in C^{k,l}(\mathbb{R}^d \times \Theta; \mathbb{R}^q)$  that is of polynomial growth uniformly in  $\theta \in \Theta$ :  $\sup_{\theta \in \Theta} |\nabla_x^\alpha \nabla_\theta^\beta f(x, \theta)| \lesssim (1 + |x|)^C$  for any  $\alpha \leq k$  and  $\beta \leq l$ .
- For a function  $g(x, \theta) : \mathbb{R}^d \times \Theta \rightarrow \mathbb{R}^q$ , we write  $g_{k-1}(\theta) := g(X_{t_{k-1}^n}, \theta)$ . Moreover, denote by  $\chi_k(\theta) := \Delta_k^n X - b_{k-1}(\theta) \Delta_n (\in \mathbb{R}^d)$ .
- All the asymptotic symbol are described under  $n \rightarrow \infty$  and  $\epsilon \rightarrow 0$  unless otherwise noted.

Using the above notation, our estimating function given in (1.7) is rewritten as

$$\Phi_{n,\epsilon}(\theta) = \epsilon^{-2} \Delta_n^{-1} \sum_{k=1}^n I[\chi_k^{\otimes 2}(\theta)] \mathbf{1}_{\{|\Delta_k^n X| \leq \delta_{n,\epsilon}\}},$$

where  $I$  is the  $d \times d$  identity matrix.

We make the following assumptions on the model (1.1):

**A1**  $|b(x, \theta) - b(y, \theta)| \lesssim |x - y|$  for each  $x, y \in \mathbb{R}^d$  and  $\theta \in \Theta$ .

Under this assumption, the ordinary differential equation

$$dX_t^0 = b(X_t^0, \theta_0) dt, \quad X_0^0 = x,$$

has the unique solution  $X^0 = (X_t^0)_{t \geq 0}$ .

**A2**  $b \in C_\uparrow^{2,3}(\mathbb{R}^d \times \Theta; \mathbb{R}^d)$ .

**A3**  $\theta \neq \theta_0 \Leftrightarrow b(X_t^0, \theta) \neq b(X_t^0, \theta_0)$  for at least one value of  $t \in [0, 1]$ .

**A4**  $I(\theta_0) := \int_0^1 \nabla_\theta b(X_t^0, \theta_0)^\top \nabla_\theta b(X_t^0, \theta_0) dt (\in \mathbb{R}^p \otimes \mathbb{R}^p)$  is positive definite,

We further make the following conditions for the limiting process  $Q$  of  $Q^\epsilon$ .

**Q1** $[\gamma]$  There exists some  $\gamma > 0$  such that, for any  $k = 1, \dots, n$ ,

$$\mathbb{P} \left\{ \sup_{t \in (t_{k-1}^n, t_k^n]} |Q_t - Q_{t_{k-1}^n}| > \Delta_n^\gamma \middle| \mathcal{F}_{t_{k-1}^n} \right\} = o_p(1).$$

**Q2[q]** For  $q > 0$  and processes  $A$  and  $M$  given in (1.3), the total variation of  $A$ , say  $TV(A) := \int_0^1 |dA_t|$ , and the quadratic variation  $[M, M]$  satisfy that

$$\mathbb{E} [TV(A)^q] + \mathbb{E} \left[ [M, M]_1^{q/2} \right] < \infty.$$

Although the condition Q1 seems *ad hoc*, we can give some easy-to-check conditions in some important cases where, e.g.,  $Q$  is a Lévy process as well as Q2; see Section 2.4 for details.

## 2.2 Asymptotic behavior of threshold-type estimators

**Theorem 1.** Suppose A1–A3, Q1[ $\gamma$ ], and that a sequence  $\{\delta_{n,\epsilon}\}$  satisfies that

$$\delta_{n,\epsilon} \Delta_n^{-1} \rightarrow \infty, \quad \epsilon \Delta_n^\gamma \delta_{n,\epsilon}^{-1} = O(1). \quad (2.1)$$

Then

$$\hat{\theta}_{n,\epsilon} \xrightarrow{\mathbb{P}} \theta_0.$$

**Remark 1.** Condition (2.1) ensures a kind of “negligibility”:

$$\mathbb{P} \left( |\Delta_k^n X| > \delta_{n,\epsilon} | \mathcal{F}_{t_{k-1}^n} \right) \rightarrow 0,$$

which makes  $\hat{\theta}_{n,\epsilon}$  asymptotically equivalent to  $\hat{\theta}_{n,\epsilon}^{LSE}$ ; see (5.13) in the proof.

**Theorem 2.** Suppose the same assumptions as in Theorem 1, and further A4. Then

$$\epsilon^{-1}(\hat{\theta}_{n,\epsilon} - \theta_0) \xrightarrow{\mathbb{P}} \zeta := I^{-1}(\theta_0) \int_0^1 \nabla_{\theta} b(X_t^0, \theta_0) [dQ_t],$$

**Remark 2.** In Long et al. [20], Theorem 2.2 assumes that

$$n\epsilon \rightarrow \infty.$$

Note that this condition is implied by (2.1) which implies that

$$n\delta_{n,\epsilon} \rightarrow \infty, \quad \epsilon/\delta_{n,\epsilon} \rightarrow \infty.$$

**Theorem 3.** Suppose the same assumptions as in Theorem 2, and that Q2[M] holds true for any  $M > 0$ . Then, it follows for every continuous function  $f : \mathbb{R}^p \rightarrow \mathbb{R}$ , of polynomial growth that

$$\mathbb{E} \left[ f \left( \epsilon^{-1}(\hat{\theta}_{n,\epsilon} - \theta_0) \right) \right] \rightarrow \mathbb{E}[f(\zeta)],$$

where  $\zeta$  is given in Theorem 2.

## 2.3 Examples and remarks

### 2.3.1 $\alpha$ -stable noise

Suppose  $d = 1$ , and that  $Q^\epsilon = \sigma B + \rho_\epsilon S$ , where  $\rho_\epsilon \rightarrow \rho \in [0, \infty)$ ,  $\sigma \geq 0$  is a constant,  $B$  is a Brownian motion, and  $S$  is a standard  $\alpha$ -stable motion with stability index  $\alpha \in (0, 2)$  and the skewness parameter  $\beta \in [-1, 1]$ , which is denoted by  $S_\alpha(1, \beta, 0)$ ; see, e.g., Cont and Tankov [4], page 94 for this notation. In this case, the limiting variable  $\zeta$  in Theorem 2 becomes

$$\zeta = I^{-1}(\theta_0) \{ \sigma D_2 \cdot Z + \rho D_\alpha \cdot S_\alpha(1, \beta, 0) \},$$

where  $D_\alpha = \left( \int_0^1 \{ \partial_\theta b(X_t^0, \theta_0) \}^\alpha dt \right)^{1/\alpha}$  and  $Z$  is a standard Gaussian variable. Hence if  $\rho = 0$  then the estimator is asymptotically normal; see also Long et al. [20] for more a specific SDE.

### 2.3.2 Markovian noise

Let us consider the case where

$$dQ_t^\epsilon = \epsilon_1 a(X_t^\epsilon) dW_t + \epsilon_2 c(X_{t-}^\epsilon) dZ_t, \quad (2.2)$$

where  $\epsilon_j \rightarrow 0$  ( $j = 1, 2$ ),  $a, c$  are some suitable functions, and  $Z$  is a pure jump Lévy process.

- Case of  $\rho_\epsilon := \epsilon_2/\epsilon_1 \rightarrow 0$ : we can reparametrize as  $\tilde{\epsilon} := \epsilon \cdot \epsilon_1$  to obtain that

$$dX_t^\epsilon = b(X_t^\epsilon, \theta_0) dt + \tilde{\epsilon} \cdot d\tilde{Q}_t^\epsilon,$$

with

$$d\tilde{Q}_t^\epsilon = a(X_t^\epsilon) dW_t + \rho_\epsilon c(X_{t-}^\epsilon) dZ_t.$$

Then  $\tilde{\epsilon}^{-1}(\hat{\theta}_{n,\epsilon} - \theta_0)$  is asymptotically normal as in Sørensen and Uchida [30]. We can verify that

$$\tilde{Q}_t^\epsilon \rightarrow Q := \int_0^t a(X_s^0) dW_s \quad a.s.,$$

uniformly in  $t \in [0, 1]$  from the fact that the stochastic integrals are continuous with respect to the *u.c.p. topology*; see Theorem II.11 by Protter [24], and the uniform convergence of  $X^\epsilon \rightarrow X^0$  on compact sets; see Theorem IX.4.21 by Jacod and Shiryaev [10].

- Case of  $\rho_\epsilon := \epsilon_2/\epsilon_1 \rightarrow \rho \in (0, \infty)$ : we can use the same model as above, and the asymptotic distribution of  $\tilde{\epsilon}^{-1}(\hat{\theta}_{n,\epsilon} - \theta_0)$  is a convolution of stochastic integrals with respect to  $W$  and  $Z$ .

- Case of  $\rho_\epsilon := \epsilon_2/\epsilon_1 \rightarrow \infty$ : we can reparametrize as  $\bar{\epsilon} := \epsilon \cdot \epsilon_2$  to obtain that

$$dX_t^\epsilon = b(X_t^\epsilon, \theta_0) dt + \bar{\epsilon} \cdot d\bar{Q}_t^\epsilon,$$

with

$$d\bar{Q}_t^\epsilon = \rho_\epsilon^{-1} a(X_t^\epsilon) dW_t + c(X_{t-}^\epsilon) dZ_t.$$

Then the asymptotic distribution of  $\bar{\epsilon}^{-1}(\hat{\theta}_{n,\epsilon} - \theta_0)$  is written by a stochastic integral with respect to  $Z$  only.

### 2.3.3 Long-term observations

An advantage of the small-noise asymptotics is that we can sometimes justify an estimation of drift parameter  $\theta$  without “ergodicity” nor uniform moment conditions such as  $\sup_{t>0} \mathbb{E}|X_t|^m < \infty$  for any  $m > 0$ , which are often essential to estimate  $\theta$  based on a ‘long-term’ observations such that  $t_n^n \rightarrow \infty$  as  $n \rightarrow \infty$ .

Consider a ‘long-term’ discretely observed model such as

$$X_t = x + \int_0^t b(X_s, \theta) ds + Q_t, \quad (2.3)$$

observed at  $0 = t_0^n < t_1^n < \dots < t_n^n =: T$  with  $T \rightarrow \infty$  as  $n \rightarrow \infty$ , and suppose that  $Q$  is an  $\alpha$ -stable process;  $Q_1 \sim S_\alpha(\sigma, \beta, 0)$  with index  $\alpha \in (1, 2]$  for simplicity. Transform the model by  $s = Tu$  and divide by  $T$  on both sides to obtain that

$$Y_s = \frac{x}{T} + \int_0^s b(T \cdot Y_u, \theta) du + \frac{1}{T} Q_{Ts}, \quad s \in [0, 1],$$

where  $Y_s = T^{-1}X_{Ts}$ . Since  $Q_{Ts} \stackrel{d}{=} T^{1/\alpha}Q_s$  by the *self-similarity* of stable processes, we can regard that  $Y$  is a (weak) solution to the following SDE:

$$Y_s = \frac{x}{T} + \int_0^s b(T \cdot Y_u, \theta) du + T^{1/\alpha-1} \tilde{Q}_s,$$

where  $\tilde{Q}_s$  is also an  $\alpha$ -stable process such that  $\tilde{Q}_1 \sim S_\alpha(\sigma, \beta, 0)$ . Suppose that  $\epsilon := T^{1/\alpha-1}$  is “small” enough ( $T$  is “large” enough), and put  $\tilde{b}(x, \theta) = b(Tx, \theta)$ , we have that

$$Y_s = \epsilon x + \int_0^s \tilde{b}(Y_u, \theta) du + \epsilon \cdot \tilde{Q}_s, \quad s \in [0, 1], \quad (2.4)$$

which is interpreted as a “small” noise SDE. This implies that when we are willing to estimate a drift parameter  $\theta$  even for possibly “non-ergodic” model under the ‘long-term’ observation as in (2.3), we can estimate it by plotting the data  $\{Y_{s_k^n} := T^{-1}X_{Ts_k^n}\}_{k=0}^n$  with  $s_k^n = t_k^n/T$  in the interval  $[0, 1]$ , and re-considering the transformed model (2.4) with known constant  $\epsilon = T^{-1}$ .

We remark that the similar argument is possible for the case where  $Q$  is given by (2.2).



## 2.4 On conditions for $Q$

Let us investigate some sufficient conditions to ensure Q1 and Q2 when  $Q$  is specified.

An important case is when  $Q$  is a Lévy process such that the *characteristic exponent*:  $\psi(u) := \log \mathbb{E} [\exp(iu^\top Q_1)]$ , is given by

$$\psi(u) = ib^\top u - \frac{\sigma^2}{2} u^\top u + \int_{\mathbb{R}^d} \left( e^{iu^\top z} - 1 - \frac{iu^\top z}{1 + |z|^2} \right) \nu(dz), \quad u \in \mathbb{R}^d, \quad (2.5)$$

where  $b \in \mathbb{R}^d$ ,  $\nu$  is the Lévy measure with  $\nu(\{0\}) = 0$  and  $\int_{|z| \leq 1} |z|^2 \nu(dz) < \infty$ .

As a simple case such that  $Q$  is a Wiener process, where  $b = 0, \nu \equiv 0$ , it follows from the property of the stationary, independent increments that

$$\begin{aligned} \mathbb{P} \left\{ \sup_{t \in (t_{k-1}^n, t_k^n]} |W_t - W_{t_{k-1}^n}| > \Delta_n^\gamma \middle| \mathcal{F}_{t_{k-1}^n} \right\} &= \mathbb{P} \left\{ \sup_{t \in (0, \Delta_n]} |W_t| > \Delta_n^\gamma \right\} \\ &= 2 \left( 1 - \Phi(\Delta_n^\gamma / \sqrt{\Delta_n}) \right) \rightarrow 0, \end{aligned}$$

for any  $\gamma \in (0, 1/2)$ , where  $\Phi$  is a standard normal distribution function; for the last equality, see, e.g., Doob [5], or Boukai [3], etc.

When  $Q$  is a stable process, we have the following result.

**Proposition 1.** *When  $Q$  is a symmetric  $\alpha$ -stable process with  $\alpha \in (1, 2)$ . Then  $Q1[\gamma]$  holds true for any  $\gamma \in (0, \alpha^{-1})$ .*

*Proof.* Due to the maximal inequality (3.5) in Joulin [11], we have

$$\mathbb{P} \left\{ \sup_{t \in (t_{k-1}^n, t_k^n]} |Q_t| > \Delta_n^\gamma \middle| \mathcal{F}_{t_{k-1}^n} \right\} = \mathbb{P} \left\{ \sup_{t \in (0, \Delta_n]} |Q_t| > \Delta_n^\gamma \right\} = O(\Delta_n^{1-\gamma\alpha}) \rightarrow 0$$

as  $n \rightarrow \infty$ . □

We can also present some sufficient conditions to  $Q1[\gamma]$  when a process  $Q$  is a more general Lévy process: let

$$\begin{aligned} h(x) &:= \int_{|z| > x} \nu(dz) + x^{-2} \int_{|z| \leq x} |z|^2 \nu(dz) \\ &\quad + x^{-1} \left| b + \int_{|z| \leq x} \frac{z|z|^2}{1 + |z|^2} \nu(dz) - \int_{|z| > x} \frac{z}{1 + |z|^2} \nu(dz) \right|. \end{aligned}$$

**Proposition 2.** *Suppose that  $Q$  is a Lévy process with characteristic (2.5), and that there exists a constant*

$$\beta := \inf \left\{ \eta > 0 : \limsup_{x \rightarrow 0} x^\eta h(x) = 0 \right\}. \quad (2.6)$$

Then the condition  $Q1[\gamma]$  holds true for any  $\gamma \in (0, \gamma_0)$ , where

$$\gamma_0 = \begin{cases} \beta^{-1} & (\sigma = 0) \\ (\beta \vee 2)^{-1} & (\sigma \neq 0) \end{cases}.$$

We interpret that  $1/0 = \infty$ .

*Proof.* Let  $Q_t = \sigma W_t + Z_t$ , where  $W$  is a Wiener process and  $Z$  is a pure jump Lévy process with characteristic

$$\mathbb{E} \left[ \exp(iu^\top Z_1) \right] = ib^\top u + \int_{\mathbb{R}^d} \left( e^{iu^\top z} - 1 - \frac{iu^\top z}{1 + |z|^2} \right) \nu(dz), \quad u \in \mathbb{R}^d.$$

By the independent, stationary increments property for Lévy process  $Q$ , we see that

$$\begin{aligned} R_n &:= \mathbb{P} \left\{ \sup_{t \in (t_{k-1}^n, t_k^n]} |Q_t - Q_{t_{k-1}^n}| > \Delta_n^\gamma \middle| \mathcal{F}_{t_{k-1}^n} \right\} \\ &\lesssim \mathbb{P} \left\{ \sup_{t \in (0, \Delta_n]} |\sigma W_t| > \Delta_n^\gamma/2 \right\} + \mathbb{P} \left\{ \sup_{t \in (0, \Delta_n]} |Z_t| > \Delta_n^\gamma/2 \right\} =: R_n^{(1)} + R_n^{(2)}. \end{aligned}$$

Note that  $R_n^{(1)} \rightarrow 0$  for any  $\alpha \in (0, 1/2)$  as  $\sigma \neq 0$ , and that  $R_n^{(1)} \equiv 0$  if  $\sigma = 0$ . Moreover, according to Pruitt [25], (3.2), it follows that

$$R_t^{(2)} \lesssim \Delta_n h(\Delta_n^\gamma/2) = 2x_n^{1/\gamma} h(x_n), \quad x_n := \Delta_n^\gamma/2.$$

Hence, if  $1/\gamma > \beta$ , equivalently  $\gamma \in (0, \beta^{-1})$ , then  $R_n^{(2)} \rightarrow 0$  by the definition of  $\beta \geq 0$ . Then the result follows.  $\square$

**Corollary 1.** Suppose that  $\int_{|z| \leq 1} |z| \nu(dz) < \infty$  and that the Lévy characteristic  $\psi$  is given by

$$\psi(u) = \frac{\sigma^2}{2} u^\top u + \int_{\mathbb{R}^d} \left( e^{iu^\top z} - 1 \right) \nu(dz).$$

Then  $\beta$  in Proposition 2, (2.6) is given by the Blumenthal-Gettoor index:

$$\beta = \inf \left\{ \eta > 0 : \int_{|z| \leq 1} |z|^\eta \nu(dz) < \infty \right\} \leq 1.$$

*Proof.* Note that, under the assumption, we have

$$h(x) = \int_{|z| > x} \nu(dz) + x^{-2} \int_{|z| \leq x} |z|^2 \nu(dz) + x^{-1} \left| \int_{|z| \leq x} \frac{z|z|^2}{1 + |z|^2} \nu(dz) \right|.$$

Let  $\beta_0$  be the Blumenthal-Gettoor index:

$$\beta_0 = \inf \left\{ \eta \in [0, 2] : \int_{|z| \leq 1} |z|^\eta \nu(dz) < \infty \right\}.$$

Then we easily see by the direct computation that, for any  $\epsilon > 0$

$$\lim_{x \rightarrow 0} x^{\beta_0 + \epsilon} h(x) = 0, \quad \liminf_{x \rightarrow 0} x^{\beta_0 - \epsilon} h(x) = \infty.$$

Indeed, for  $x \in (0, 1)$ , we have

$$\begin{aligned} |x^{\beta_0 + \epsilon} h(x)| &\leq |x|^{\beta_0 + \epsilon} \left| \int_{x < |z| \leq 1} \nu(dz) + \int_{|z| > 1} \nu(dz) \right| \\ &\quad + |x|^{\beta_0 + \epsilon - 2} \int_{|z| \leq x} |z|^2 \nu(dz) + |x|^{\beta_0 + \epsilon - 1} \int_{|z| \leq x} \frac{|z|^3}{1 + |z|^2} \nu(dz) \\ &\leq |x|^\epsilon \int_{|z| \leq 1} |z|^{\beta_0} \nu(dz) + \left| x^{\beta_0 + \epsilon} \int_{|z| > 1} \nu(dz) \right| + |x|^{\beta_0 + \epsilon - 1} \int_{|z| \leq x} |z|^{1 - \beta_0} \nu(dz) \\ &= O(|x|^\epsilon) \rightarrow 0. \end{aligned}$$

The other condition:  $\liminf_{x \rightarrow 0} x^{\beta_0 - \epsilon} h(x) = \infty$  is easier since

$$x^{\beta_0 - \epsilon} h(x) > x^{\beta_0 - \epsilon - 2} \int_{|z| \leq x} |z|^2 \nu(dz) \geq x^{-\epsilon} \int_{|z| \leq x} |z|^{\beta_0} \nu(dz) \rightarrow \infty,$$

as  $x \rightarrow 0$ . Hence we get  $\beta_0 = \beta$ . □

The condition Q2 is reduced to moment conditions for the Lévy measure when  $Q$  is a Lévy process given by (2.5).

**Proposition 3.** *Suppose that  $Q$  is a Lévy process with characteristic (2.5). Then the condition Q2[q] is equivalent to*

$$\int_{|z| > 1} |z|^q \nu(dz) < \infty.$$

*Proof.* A Lévy process  $Q$  has the following Lévy-Ito representation:

$$Q_t = at + bW_t + \int_0^t \int_{|z| \leq 1} z \tilde{N}(ds, dz) + \int_0^t \int_{|z| > 1} z N(ds, dz)$$

where  $a, b$  are constants,  $W$  is a Wiener process,  $N$  is a Poisson random measure, and  $\tilde{N}(ds, dz) = N(ds, dz) - \nu(dz)ds$ . That is,

$$A_t = at + \int_0^t \int_{|z| > 1} z N(ds, dz), \quad M_t = bW_t + \int_0^t \int_{|z| \leq 1} z \tilde{N}(ds, dz).$$

Hence, it follows from the Burkholder-Davis-Gundy inequality that

$$\mathbb{E}[TV(A)^q] \lesssim |a|^q + \mathbb{E} \left| \int_0^1 \int_{|z| > 1} |z| N(ds, dz) \right|^q$$

$$\lesssim 1 + \mathbb{E} \left| \int_0^1 \int_{|z|>1} |z| \tilde{N}(ds, dz) \right|^q + \int_{|z|>1} |z|^q \nu(dz).$$

According to the argument as in Bichteler and Jacod [2]; see also the proofs of lemma 4.1 and Proposition 3.1 by Shimizu and Yoshida [28], we see that

$$\mathbb{E} \left| \int_0^1 \int_{|z|>1} |z| \tilde{N}(ds, dz) \right|^q \lesssim \int_{|z|>1} |z|^q \nu(dz).$$

Hence

$$\mathbb{E}[TV(A)^q] \lesssim \int_{|z|>1} |z|^q \nu(dz) < \infty.$$

By the similar argument, it is easy to see from Hölder's inequality that, for any  $m > 1$ ,

$$\mathbb{E}[|M, M|^{q/2}] \leq \left( \mathbb{E}[|M, M|^{mq/2}] \right)^{1/m} \lesssim 1 + \left( \int_{|z|\leq 1} |z|^{mq/2} \nu(dz) \right)^{1/m} < \infty,$$

if we take  $mq/2 \geq 2$ . This completes the proof.  $\square$

### 3 Numerical studies

#### 3.1 2-dim model

We consider the following 2-dimensional Lévy driven SDE:

$$b(x, \theta) = \left( \sqrt{\theta_1 + x_1^2 + x_2^2}, -\frac{\theta_2 x_2}{\sqrt{1 + x_1^2 + x_2^2}} \right)^\top, \quad Q_t = \begin{pmatrix} V_t^{\kappa, \xi} + B_t \\ S_t^\alpha \end{pmatrix}, \quad (3.1)$$

where  $B$  is a standard Brownian motion,  $S^\alpha$  is a standard symmetric  $\alpha$ -stable process  $S_\alpha(1, 0, 0)$ , and  $V^{\delta, \gamma}$  is a variance gamma process with Lévy density

$$p_V(z) = \frac{\kappa}{|z|} e^{-\xi|z|}, \quad z \in \mathbb{R}, \quad \kappa, \xi > 0,$$

which is obtained by Brownian subordination with a gamma process  $G_t \sim \Gamma(\text{shape} = ct, \text{scale} = 1/\lambda)$  ( $c, \lambda > 0$ ) as follows:

$$V_t^{\kappa, \xi} = \sigma W_{G_t} \quad (\sigma > 0), \quad \kappa = \frac{\lambda^2}{c}, \quad \xi = \frac{\sqrt{2\kappa}}{\sigma}$$

where  $W$  is the standard Brownian motion independent of  $G$ ; see, e.g., Cont and Tankov [4] for details. Assume that  $W, S^\alpha$  and  $V^{\delta, \gamma}$  are independent each other.

In the sequel, we set values of parameters as

$$(X_0^{(1)}, X_0^{(2)}) = (1, 1), \quad (\theta_1, \theta_2) = (2, 1), \quad (\kappa, \xi, \alpha) = (5, 3, 3/2).$$

A sample path is given in Figure 5. Then both  $X^{(1)}$  and  $X^{(2)}$  are unbounded variation jump-processes with finite activity of jumps for  $X^{(1)}$  and infinite for  $X^{(2)}$ . We will compare our threshold-type estimator to the LSE by Long et al. [20].

### 3.2 LSE and threshold estimator

Note that the LSE  $\hat{\theta}^{LSE} = (\hat{\theta}_{1,n,\epsilon}^{LSE}, \hat{\theta}_{2,n,\epsilon}^{LSE})$  is a solution to

$$\sum_{k=1}^n \frac{\Delta_k^n X^{(1)}}{\sqrt{\hat{\theta}_{1,n,\epsilon}^{LSE} + (X_{t_{k-1}^n}^{(1)})^2 + (X_{t_{k-1}^n}^{(2)})^2}} = 1; \quad \hat{\theta}_{2,n,\epsilon}^{LSE} = - \frac{\sum_{k=1}^n \frac{(\Delta_k^n X^{(2)}) X_{t_{k-1}^n}^{(2)}}{\sqrt{1 + (X_{t_{k-1}^n}^{(1)})^2 + (X_{t_{k-1}^n}^{(2)})^2}}}{n^{-1} \sum_{k=1}^n \frac{(X_{t_{k-1}^n}^{(2)})^2}{1 + (X_{t_{k-1}^n}^{(1)})^2 + (X_{t_{k-1}^n}^{(2)})^2}}.$$

and that our estimator  $\hat{\theta} = (\hat{\theta}_{1,n,\epsilon}, \hat{\theta}_{2,n,\epsilon})$  is a solution to

$$\sum_{k=1}^n \frac{\Delta_k^n X^{(1)}}{\sqrt{\hat{\theta}_{1,n,\epsilon} + (X_{t_{k-1}^n}^{(1)})^2 + (X_{t_{k-1}^n}^{(2)})^2}} \mathbf{1}_{\{|\Delta_k^n X| \leq \delta_{n,\epsilon}\}} = \frac{1}{n} \sum_{k=1}^n \mathbf{1}_{\{|\Delta_k^n X| \leq \delta_{n,\epsilon}\}}$$

$$\hat{\theta}_{2,n,\epsilon} = - \frac{\sum_{k=1}^n \frac{(\Delta_k^n X^{(2)}) X_{t_{k-1}^n}^{(2)}}{\sqrt{1 + (X_{t_{k-1}^n}^{(1)})^2 + (X_{t_{k-1}^n}^{(2)})^2}} \mathbf{1}_{\{|\Delta_k^n X| \leq \delta_{n,\epsilon}\}}}{n^{-1} \sum_{k=1}^n \frac{(X_{t_{k-1}^n}^{(2)})^2}{1 + (X_{t_{k-1}^n}^{(1)})^2 + (X_{t_{k-1}^n}^{(2)})^2} \mathbf{1}_{\{|\Delta_k^n X| \leq \delta_{n,\epsilon}\}}}.$$

In the simulations, we tried  $n = 1000, 3000, 5000$  and  $\epsilon = (0, 4, 0.3, 0.05)$ . The results for the LSE given in (1.4) are in Table 1, and those for the threshold-type estimator in (1.6) are in 2 and 3, in which the threshold  $\delta_{n,\epsilon} = \epsilon/5$  and  $\epsilon/10$  are used, respectively. The simulations are iterated 10000 times, and mean and standard deviation (s.d.) of estimators are given in those tables. In Tables 2 and 3, the values of  $(n\delta_{n,\epsilon}, \delta_{n,\epsilon}\epsilon^{-1}n^{1/4})$  are also included. Due to the assumptions in Theorems 1–3, those values goes to  $(\infty, 0)$  as  $n \rightarrow \infty, \epsilon \rightarrow 0$  in this example.

### 3.3 Discussion

From numerical results, we can observe that the threshold-type estimator improves the accuracy of estimation in the sense of the standard deviation compared with the usual LSE by Long et al. [20]. Especially, the improvement for  $\theta_2$  is more drastic than that for  $\theta_1$ . This would be because  $\theta_2$  is the mean-reverting parameter for  $X^{(2)}$ , a *stable process* which has much more frequent and larger jumps than those of  $X^{(1)}$ . In order to make the estimation of  $\theta_1$  more accurate, we need to make  $\epsilon$  smaller rather than  $n$  larger.

We should note that the result for  $(n, \epsilon, \delta_{n,\epsilon}) = (1000, 0.05, \epsilon/10)$ , where estimation has a large bias. A reason for this ‘bad’ estimation would be that the asymptotic theory does not work well because  $n\delta_{n,\epsilon} = 5$  is ‘small’ although that should be large enough theoretically. From many simulations omitted here, we see that, at least, “ $n\delta_{n,\epsilon} \geq 10$ ”

would be needed for estimation with ‘small’ bias; see also, e.g., the case where  $(n, \epsilon, \delta_{n,\epsilon}) = (1000, 0.05, \epsilon/5)$ , which returns better estimation.

Although there remains a problem to choose  $\delta_{n,\epsilon}$  in practice, we can use the method proposed by Shimizu [27] if the parameters in noise process is known or estimable.

Finally, we observe normal QQ-plots for normalized estimators  $\epsilon^{-1}(\hat{\theta}_{1,n,\epsilon} - \theta_1)$  and  $\epsilon^{-1}(\hat{\theta}_{2,n,\epsilon} - \theta_2)$  in the case where  $n = 5000$ ,  $\epsilon = 0.1$  and  $\delta = \epsilon/5$ ; see Figures 3 and 4. According to the results, the estimators with  $\delta = \epsilon/5$  seem asymptotically normal although Theorem 2 does not necessarily say that. For your reference, see Figures 1 and 2 that are the normal QQ-plot for the LSE without filter proposed by Long et al. [20]. The figure shows that the usual LSE is not necessarily asymptotically normal as the theory saying. Therefore, the asymptotic normality-like phenomena would be due to the filter effects.

We could understand those results intuitively as follows: cutting large jumps from a process with infinite activity jumps, the remaining small jumps will behave as a Brownian motion. For example, suppose that the driving noise  $Q$  be a Lévy process of infinite activity with the Lévy measure  $\nu$ , and put

$$Q_t^{(\delta)} := \int_0^t \int_{|z| \leq \delta} z \tilde{N}(ds, dz), \quad \delta > 0,$$

where  $\tilde{N}$  is a compensated Poisson random measure as given in the proof of Proposition 3. Then, according to Asmussen and Rosinski [1], it follows for  $\sigma^2(\delta) := \int_{|z| \leq \delta} |z|^2 \nu(dz)$  that

$$\sigma(\delta)^{-1} Q^{(\delta)} \xrightarrow{\mathcal{D}} B,$$

in  $\mathbb{D}[0, 1]$ -space equipped with the sup-norm under a certain assumption, where  $B$  is a standard Brownian motion. Therefore, by an appropriate norming of estimator, the limit in Theorem 2 might be an integral with respect to the process that is close to a Brownian motion. This consideration indicates that the LSE with filter can be “approximately” asymptotically normal by letting  $\delta_{n,\epsilon}$  converge to zero in a suitable rate, which is a great advantage when we would like to make confidence intervals or do statistical testing. In the next subsection, we shall try to understand this phenomenon theoretically.

## 4 Could a filtered LSE be asymptotically normal?

In this section, we assume that  $Q^\epsilon \equiv Q$  that is a Lévy process with Lévy measure  $\nu$ . We consider the following two cases for  $Q$  given by (2.5):

- Finite activity case:  $\int_{|z| \leq 1} \nu(dz) < \infty$  and  $\sigma^2 > 0$ ;
- Infinite activity case:  $\int_{|z| \leq 1} \nu(dz) = \infty$  (possibly,  $\sigma^2 = 0$ ).

The former case, it would be possible to show the asymptotic normality of the filtered LSE by separating the increments  $\Delta_k^n X$ 's with or without jumps as in Shimizu and Yoshida [28]. However, when  $\int_{|z| \leq 1} \nu(dz) = \infty$ , it is known that the filter  $\mathbf{1}_{\{|\Delta_k^n X| \leq \delta_{n,\epsilon}\}}$  is not enough to separate  $\Delta_k^n X$ 's with or without jumps; see Shimizu [26], Lemma 3.3 and some comments on that.

In this section, in order to understand essentially why the filtered LSE seems to be asymptotically normal, we shall consider the following *ad hoc* situation.

**[Assumption]** *All the jumps of  $Q$  are observed.*

Under this assumption, the following contrast function does make sense.

$$\tilde{\Phi}_{n,\epsilon,\delta}(\theta) = \epsilon^{-2} \Delta_n^{-1} \sum_{k=1}^n |\Delta_k^n X - b(X_{t_{k-1}^n}, \theta) \cdot \Delta_n|^2 \mathbf{1}_{\{\|\Delta Q^\epsilon\|_k^* \leq \delta\}},$$

where  $\delta > 0$  is a constant, and

$$\|Q^\epsilon\|_k^* = \sup_{t \in (t_k, t_{k+1}]} |\Delta Q_t^\epsilon|, \quad \Delta Q_t^\epsilon = \epsilon \cdot (Q_t - Q_{t-}).$$

Under our assumption, we can specify if  $\mathbf{1}_{\{\|\Delta Q^\epsilon\|_k^* \leq \delta\}} = 1$  or 0, and we can define the estimator of  $\theta$  as the minimum contrast estimator:

$$\tilde{\theta}_{n,\epsilon,\delta} = \arg \min_{\theta \in \Theta} \tilde{\Phi}_{n,\epsilon,\delta}(\theta).$$

Hereafter, we further use the following notation:

- For a  $\kappa > 0$ ,

$$\sigma^2(\kappa) := \int_{|z| \leq \kappa} |z|^2 \nu(dz); \quad \lambda(\kappa) := \int_{|z| > \kappa} \nu(dz),$$

where  $\nu$  is a Lévy measure of  $Q$ .

- All the asymptotic symbols are used under  $\delta, \epsilon \rightarrow 0$  and  $n \rightarrow \infty$ .

#### 4.1 Finite activity case

Suppose that  $\int_{|z| \leq 1} \nu(dz) < \infty$  and  $\sigma^2 > 0$ , which implies that  $Q$  is written as

$$Q_t = \sigma W_t + \sum_{i=1}^{N_t} Y_i, \tag{4.1}$$

where  $N$  is a Poisson process with intensity  $\lambda := \int_{\mathbb{R}} \nu(dz)$  and  $\{Y_i\}_{i=1,2,\dots}$  is an i.i.d. sequence with distribution  $\lambda^{-1}\nu$ . In this special case, we have the following result by taking  $\delta_{n,\epsilon} \downarrow 0$  faster than the speed of  $\epsilon$  that is a “magnitude of jumps”.

**Theorem 4.** Suppose that  $Q$  is given by (4.1), and that A1–A4 hold true. Moreover, suppose that

$$\delta/\epsilon \rightarrow 0. \quad (4.2)$$

Then

$$\epsilon^{-1}(\tilde{\theta}_{n,\epsilon,\delta} - \theta_0) \xrightarrow{\mathbb{P}} I^{-1}(\theta_0) \int_0^1 \nabla_{\theta} b(X_t^0, \theta_0) [dW_t],$$

Hence  $\tilde{\theta}_{n,\epsilon,\delta}$  is asymptotically normal.

## 4.2 Infinite activity case

**Theorem 5.** Suppose that  $Q^\epsilon \equiv Q$  is a Lévy process with  $\int_{|z|<1} \nu(dz) = \infty$ , and A1–A4, Q1[ $\gamma$ ], and that Q2[ $q$ ] hold true for any  $q > 0$ . Moreover suppose that

$$\frac{\lambda(\delta/\epsilon)}{n \log n} \rightarrow c \in (0, 1), \quad n\epsilon \Delta_n^\gamma \rightarrow 0, \quad (4.3)$$

and that there exists a constant  $\rho \in (0, 1)$  such that

$$\sigma^\rho(\delta/\epsilon) \log n \rightarrow \infty; \quad (4.4)$$

$$n\epsilon \cdot \sigma(\delta/\epsilon) \rightarrow \infty, \quad (4.5)$$

and that, for each  $\kappa > 0$ ,

$$\sigma(\kappa\sigma(\delta/\epsilon) \wedge \delta/\epsilon) \sim \sigma(\delta/\epsilon), \quad (4.6)$$

Then there exists a  $d$ -dimensional Brownian motion  $B$ , independent of  $X_0 = x$ , such that the following weak convergence holds true:

$$(\sigma(\delta/\epsilon)\epsilon)^{-1} (\tilde{\theta}_{n,\epsilon,\delta} - \theta_0) \xrightarrow{\mathcal{D}} I^{-1}(\theta_0) \int_0^1 \nabla_{\theta} b(X_t^0, \theta_0) [dB_t],$$

Hence  $\tilde{\theta}_{n,\epsilon,\delta}$  is asymptotically normal.

We shall give a concrete example that satisfies the above situation.

**Example 1.** Consider the case where  $Q$  is a symmetric  $\alpha$ -stable process. The Lévy measure  $\nu$  is given by

$$\nu(ds) = \frac{C}{|z|^{1+\alpha}} dz, \quad \alpha \in (1, 2),$$

where  $C > 0$  is a constant. This is an infinite-activity model with  $\int |z| \wedge 1 \nu(dz) = \infty$ . In this case, the assumption Q1[ $\gamma$ ] holds if  $\gamma \in (0, \alpha^{-1})$  from Proposition 1. If we take  $\epsilon = O(n^{-\beta})$  for some  $\beta > 0$ , then we easily see that the asymptotic conditions in Theorem 5 hold true if

$$1 - \gamma < \beta < \frac{3}{2} - \frac{1}{\alpha}$$



while

$$\delta/\epsilon = O\left((n \log n)^{-1/\alpha}\right) \rightarrow 0.$$

In a model used in Section 3.1, this is hold if

$$\frac{1}{3} < \beta < \frac{5}{6}$$

**Remark 3.** The assumption (4.6) is to approximate a component of compensated small jumps by a Wiener process; see Theorem 2 by Asmussen and Rosinski [1]. According to Proposition 2.1 in [1], a simple condition that

$$(\delta/\epsilon)^{-1} \sigma(\delta/\epsilon) \rightarrow \infty \tag{4.7}$$

is sufficient for (4.6). Moreover, note that (4.6) requires a high jump-activity, which excludes cases where  $Q$  is a compound Poisson process or a gamma process; see some Examples 2.2–2.4 in [1]. Therefore it is assumed in (4.4) that

$$\lambda(\delta/\epsilon) \rightarrow \infty \quad (\delta \rightarrow 0).$$

**Remark 4.** Although this is the result under an ideal situation that all the infinite activity jumps are observable, we can imagine that such a phenomenon ‘approximately’ occurs in simulations presented in Figures 3 and 4. That is, when  $\Delta_n$  and  $\delta_{n,\epsilon}$  are sufficiently ‘small’, a filter  $\mathbf{1}_{\{|\Delta_k^n X| \leq \delta_{n,\epsilon}\}}$  can successfully cut jumps whose sizes are larger than  $\delta_{n,\epsilon}$ . However, it would be hard to show the similar result when the observations are completely discrete as in our original setting because the filter  $\mathbf{1}_{\{|\Delta_k^n X| \leq \delta_{n,\epsilon}\}}$  can not exactly exclude an increment that includes the jumps whose sizes are larger than  $\delta_{n,\epsilon}$ ; see also Shimizu [26] about the filter in infinite activity cases. The complete analysis in the discretely observed cases would be an important work in the future.

## 5 Proofs

### 5.1 Preliminary lemmas

We shall first establish some preliminary lemmas to show the main theorems later.

To evaluate functions of discrete samples, we use the following notation as in Long et al. [20]:

- Discretized process:  $Y_t^{n,\epsilon} := X_{[nt]/n}$  for  $t \geq 0$  and  $n \in \mathbb{N}$ , where  $[x]$  stands for the integer part of  $x \in \mathbb{R}$ .
- For each  $\epsilon > 0$ ,  $Q^\epsilon = A_t^\epsilon + M_t^\epsilon$ , where  $A^\epsilon$  is a process with finite variation, and  $M^\epsilon$  is an  $\mathcal{F}_t$ -local martingale with  $A_0^\epsilon = M_0^\epsilon = 0$  a.s.

$\epsilon = 0.4$	$n = 1000$	$n = 3000$	$n = 5000$	True
$\hat{\theta}_{1,n,\epsilon}^{LSE}$	2.54872	2.53364	2.78911	2.0
(s.d.)	(2.4370)	(2.4489)	(2.5602)	
$\hat{\theta}_{2,n,\epsilon}^{LSE}$	1.72415	1.79755	1.78645	1.0
(s.d.)	(3.5426)	(4.2814)	(2.9579)	

$\epsilon = 0.3$	$n = 1000$	$n = 3000$	$n = 5000$	True
$\hat{\theta}_{1,n,\epsilon}^{LSE}$	2.31618	2.29381	2.34757	2.0
(s.d.)	(1.8248)	(1.7926)	(1.7429)	
$\hat{\theta}_{2,n,\epsilon}^{LSE}$	1.50664	1.5275	1.53632	1.0
(s.d.)	(2.8685)	(2.7667)	(2.8160)	

$\epsilon = 0.05$	$n = 1000$	$n = 3000$	$n = 5000$	True
$\hat{\theta}_{1,n,\epsilon}^{LSE}$	2.00599	2.01071	2.01002	2.0
(s.d.)	(0.2951)	(0.2913)	(0.2938)	
$\hat{\theta}_{2,n,\epsilon}^{LSE}$	1.05963	1.04438	1.06135	1.0
(s.d.)	(1.3026)	(0.6773)	(0.7344)	

Table 1: These are results for the LSE (without filter) based on Long et al. [20]. We find that the standard deviation (s.d.) are large, especially, for  $\hat{\theta}_{2,n,\epsilon}$ , which implies an *unstability* of estimation. We would like to improve the stability by using a ‘filter’.

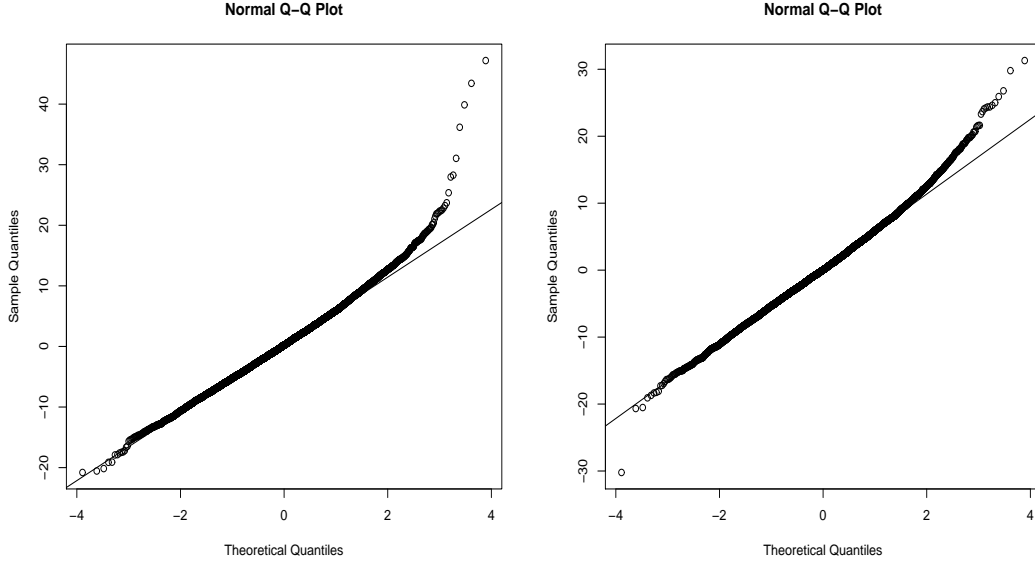


Figure 1: Normal QQ-plots for  $\widehat{\theta^{LSE}_{1,n,\epsilon}}$ . Left:  $\epsilon = 0.1$ , Right:  $\epsilon = 0.05$ . The results show that the right tail is especially heavier than that of normal distribution.

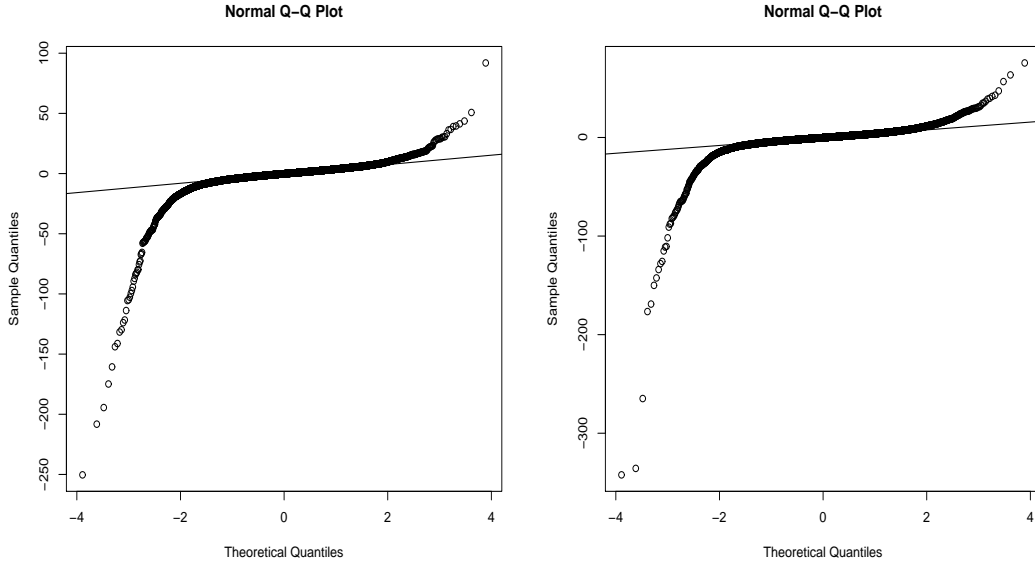


Figure 2: Normal QQ-plots for  $\widehat{\theta^{LSE}_{2,n,\epsilon}}$ . Left:  $\epsilon = 0.1$ , Right:  $\epsilon = 0.05$ . The results show that the both of tails are heavier than those of normal distribution.

$\epsilon = 0.4$	$n = 1000$	$n = 3000$	$n = 5000$	True
$\hat{\theta}_{1,n,\epsilon}$	2.46071	2.45564	2.43215	2.0
(s.d.)	(2.1968)	(2.1777)	(2.1860)	
$\hat{\theta}_{2,n,\epsilon}$	1.14289	1.15283	1.16222	1.0
(s.d.)	(0.8721)	(0.88426)	(0.8856)	
$(n\delta_{n,\epsilon}, \delta_{n,\epsilon}\epsilon^{-1}n^{1/4})$	(80,1.12)	(240,1.48)	(400,1.68)	$(\infty, 0)$

$\epsilon = 0.3$	$n = 1000$	$n = 3000$	$n = 5000$	True
$\hat{\theta}_{1,n,\epsilon}$	2.25007	2.25972	2.25829	2.0
(s.d.)	(1.6149)	(1.6121)	(1.6249)	
$\hat{\theta}_{2,n,\epsilon}$	1.08105	1.1047	1.09448	1.0
(s.d.)	(0.6498)	(0.6489)	(0.6563)	
$(n\delta_{n,\epsilon}, \delta_{n,\epsilon}\epsilon^{-1}n^{1/4})$	(60,1.12)	(180,1.48)	(300,1.68)	$(\infty, 0)$

$\epsilon = 0.05$	$n = 1000$	$n = 3000$	$n = 5000$	True
$\hat{\theta}_{1,n,\epsilon}$	2.00619	2.00827	2.00681	2.0
(s.d.)	(0.2623)	(0.2594)	(0.2652)	
$\hat{\theta}_{2,n,\epsilon}$	0.98972	0.99936	1.00039	1.0
(s.d.)	(0.10098)	(0.1031)	(0.1037)	
$(n\delta_{n,\epsilon}, \delta_{n,\epsilon}\epsilon^{-1}n^{1/4})$	(10,1.12)	(30,1.48)	(50,1.68)	$(\infty, 0)$

Table 2: Results with filter:  $\delta_{n,\epsilon} = \epsilon/5$ . Compared with the LSE, the improvement for s.d. of  $\hat{\theta}_{2,n,\epsilon}$  is drastic although the one for  $\hat{\theta}_{1,n,\epsilon}$  is less. To make the estimation of  $\theta_1$  accurate, we need to make  $\epsilon$  smaller. The values of  $(n\delta_{n,\epsilon}, \delta_{n,\epsilon}\epsilon^{-1}n^{1/4})$  seems enough to meet the asymptotic conditions such that they must tend to  $(\infty, 0)$ .

$\epsilon = 0.4$	$n = 1000$	$n = 3000$	$n = 5000$	True
$\hat{\theta}_{1,n,\epsilon}$	2.10973	2.24022	2.46723	2.0
(s.d.)	(2.0587)	(2.1736)	(2.2064)	
$\hat{\theta}_{2,n,\epsilon}$	1.06597	1.10031	1.10585	1.0
(s.d.)	(0.6997)	(0.7226)	(0.7352)	
$(n\delta_{n,\epsilon}, \delta_{n,\epsilon}\epsilon^{-1}n^{1/4})$	(40,0.56)	(120,0.74)	(150,0.84)	$(\infty, 0)$

$\epsilon = 0.3$	$n = 1000$	$n = 3000$	$n = 5000$	True
$\hat{\theta}_{1,n,\epsilon}$	1.92432	2.25282	2.24733	2.0
(s.d.)	(1.4600)	(1.6221)	(1.6021)	
$\hat{\theta}_{2,n,\epsilon}$	1.02419	1.05871	1.05869	1.0
(s.d.)	(0.5216)	(0.5453)	(0.5349)	
$(n\delta_{n,\epsilon}, \delta_{n,\epsilon}\epsilon^{-1}n^{1/4})$	(30,0.56)	(90,0.74)	(150,0.84)	$(\infty, 0)$

$\epsilon = 0.05$	$n = 1000$	$n = 3000$	$n = 5000$	True
$\hat{\theta}_{1,n,\epsilon}$	0.78959	2.00133	2.00864	2.0
(s.d.)	(0.2165)	(0.2628)	(0.2653)	
$\hat{\theta}_{2,n,\epsilon}$	0.84069	0.98834	0.99450	1.0
(s.d.)	(0.0754)	(0.0853)	(0.0865)	
$(n\delta_{n,\epsilon}, \delta_{n,\epsilon}\epsilon^{-1}n^{1/4})$	(5,0.56)	(15,0.74)	(25,0.84)	$(\infty, 0)$

Table 3: Results with filter:  $\delta_{n,\epsilon} = \epsilon/10$ . The s.d. for  $\hat{\theta}_{2,n,\epsilon}$  is smaller than those with  $\delta_{n,\epsilon}/5$  as well as those for  $\hat{\theta}_{1,n,\epsilon}$ . However, we should be careful by observing the case of  $(n, \epsilon) = (1000, 0.05)$ , where both estimators are negatively biased. This would be because  $n\delta_{n,\epsilon} = 5$  is too small to meet the corresponding asymptotic condition:  $n\delta_{n,\epsilon} \rightarrow \infty$ .

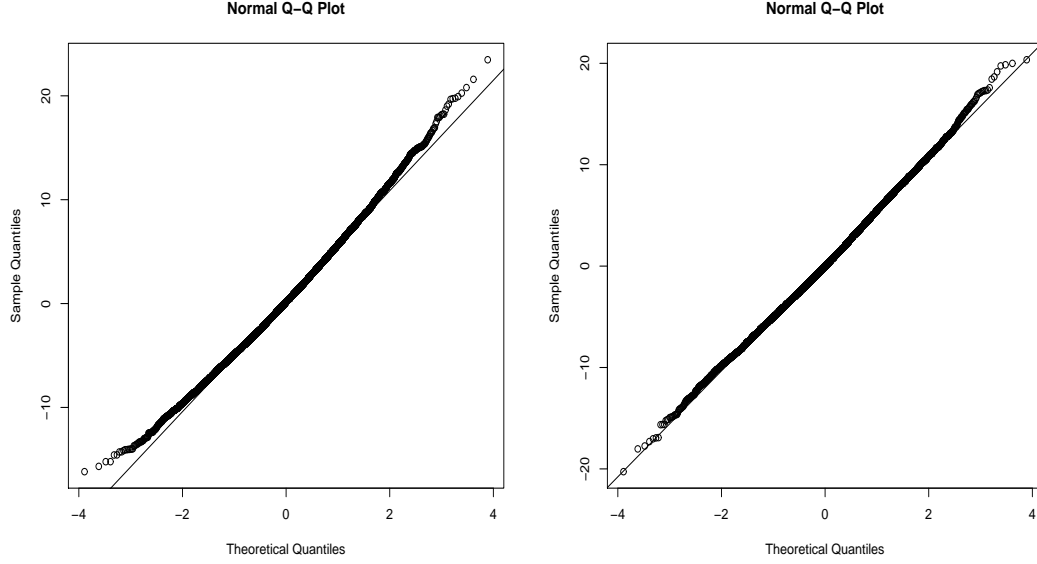


Figure 3: Normal QQ-plots for  $\hat{\theta}_{1,n,\epsilon}$ . Left:  $\epsilon = 0.1$ , Right:  $\epsilon = 0.05$ . When  $\epsilon$  is small such as  $\epsilon = 0.05$ , the distribution of  $\hat{\theta}_{1,n,\epsilon}$  seems almost Gaussian.

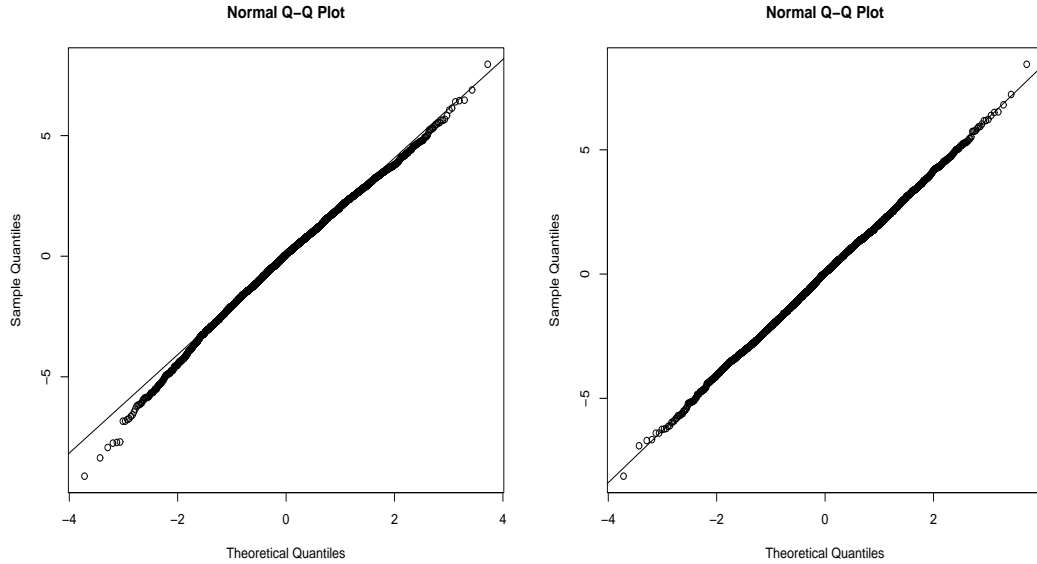


Figure 4: Normal QQ-plots for  $\hat{\theta}_{2,n,\epsilon}$ . Left:  $\epsilon = 0.1$ , Right:  $\epsilon = 0.05$ . When  $\epsilon$  is small such as  $\epsilon = 0.05$ , the distribution of  $\hat{\theta}_{1,n,\epsilon}$  seems almost Gaussian.

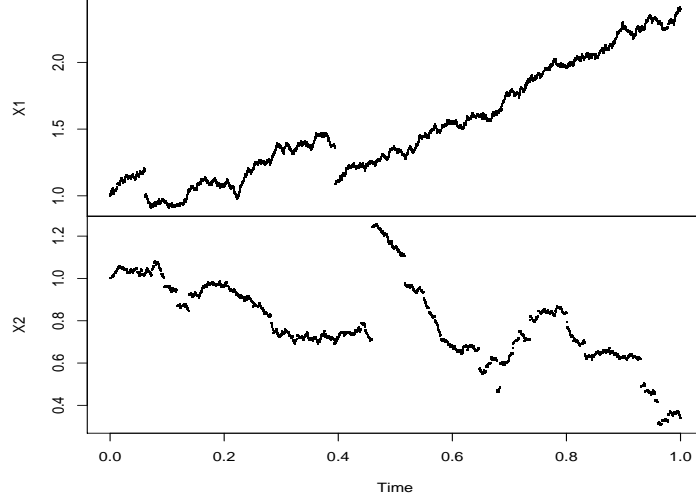


Figure 5: A sample path of Model (3.1) with  $(\theta_1, \theta_2, \kappa, \xi, \alpha) = (2, 1, 5, 3, 3/2)$  and  $\varepsilon = 0.4$ .

- A stopping time for localization: for  $m \in \mathbb{N}$ ,

$$\tau_m^{n,\epsilon} = \inf\{t \geq 0 : |X_t^0| \wedge |Y_t^{n,\epsilon}| \geq m\} \wedge T_m,$$

where  $T_m := \inf\{t \geq 0 : [M^\epsilon, M^\epsilon]_t \wedge \int_0^t |dA_s^\epsilon| \geq m\}$ . As a convention,  $\inf \emptyset = \infty$ . Hence, note that  $\lim_{m \rightarrow \infty} T_m = \infty$  almost surely.

**Lemma 1.** *Under A1, it holds that*

$$\|Y^{n,\epsilon} - X^0\|_* \xrightarrow{\mathbb{P}} 0. \quad (5.1)$$

*In addition, suppose Q2[M] for some  $M > 0$ . Then it holds that*

$$\mathbb{E} \left[ \|Y^{n,\epsilon} - X^0\|_*^M \right] = O \left( \frac{1}{n^M} + \epsilon^M \right). \quad (5.2)$$

*Proof.* From (1.2), it follows for  $\epsilon$  small enough that

$$\|Q^\epsilon\|_* \lesssim \|Q\|_* + 1. \quad (5.3)$$

Hence we can take the same argument as in the proof of Lemma 3.1, (3.1) by Long et al. [20] to obtain that

$$\|X - X^0\|_* \lesssim \epsilon \|Q^\epsilon\|_* \lesssim \epsilon (\|Q\|_* + 1) \xrightarrow{\mathbb{P}} 0, \quad (5.4)$$

since  $\|Q\|_*$  is bounded in probability. Hence the fact that  $\lfloor nt \rfloor/n \rightarrow t$  as  $n \rightarrow \infty$  yields that

$$\|Y^{n,\epsilon} - X^0\|_* \leq \|Y^{n,\epsilon} - X\|_* + \|X - X^0\|_* \xrightarrow{\mathbb{P}} 0.$$

This is the proof of (5.1).

Moreover, note from (5.4) that

$$\begin{aligned} \mathbb{E} \left[ \|Y^{n,\epsilon} - X^0\|_*^M \right] &\lesssim \mathbb{E} \left[ \sup_{t \in [0,1]} |X_{\lfloor nt \rfloor/n} - X_t|^M \right] + \mathbb{E} \left[ \|X - X^0\|_*^M \right] \\ &\lesssim \sup_{t \in [0,1]} \left( \int_t^{\lfloor nt \rfloor/n} \mathbb{E}(1 + \|X^0\|_* + \epsilon \|Q\|_*) ds \right)^M + \epsilon^M \mathbb{E} \|Q\|_*^M \\ &\lesssim \sup_{t \in [0,1]} \left[ \frac{nt - \lfloor nt \rfloor}{n} \right]^M + \epsilon^M \mathbb{E} \|A\|_*^M + \epsilon^M \mathbb{E} [M, M]^{M/2} \\ &= O \left( \frac{1}{n^M} + \epsilon^M \right), \end{aligned}$$

under Q2[M]. We used the Burkholder-Davis-Gundy inequality in the last inequality. This completes the proof.  $\square$

Form (5.2) and (5.4) in the above proof, the following corollary is obvious.

**Corollary 2.** *Suppose A1 and Q2[M] for some  $M > 0$ . Then it holds that*

$$\sup_{\epsilon > 0} \mathbb{E} \left[ (\epsilon^{-1} \|X - X^0\|_*)^M \right] < \infty. \quad (5.5)$$

In addition, if  $(n\epsilon)^{-1} = O(1)$ , then

$$\sup_{n \in \mathbb{N}, \epsilon > 0} \mathbb{E} \left[ (\epsilon^{-1} \|Y^{n,\epsilon} - X^0\|_*)^M \right] < \infty. \quad (5.6)$$

**Lemma 2.** *Under A1, it follows that*

$$\lim_{m \rightarrow \infty} \tau_m^{n,\epsilon} = \infty \quad a.s.,$$

uniformly in  $n \in \mathbb{N}$  and  $\epsilon \in [0, 1]$ .

*Proof.* Noticing (5.3), we have by the same argument as in the proof of Lemma 3.2 by Long et al. [20] that

$$\sup_{n \in \mathbb{N}, \epsilon > 0} |Y_t^{n,\epsilon}| \leq \sqrt{2} \left( |x| + \sup_{s \in [0,t]} |Q_s| + t \right) e^{Ct^2} < \infty \quad a.s.$$

for any  $t > 0$ . Therefore we have the consequence.  $\square$



**Lemma 3.** *Let  $g \in C_{\uparrow}^{1,1}(\mathbb{R}^d \times \Theta; \mathbb{R})$ . Suppose A1, A2, Q1[ $\gamma$ ], and that*

$$\delta_{n,\epsilon} \Delta_n^{-1} \rightarrow \infty, \quad \epsilon \Delta_n^\gamma \delta_{n,\epsilon}^{-1} = O(1).$$

*Then, we have*

$$\frac{1}{n} \sum_{k=1}^n g_{k-1}(\theta) \mathbf{1}_{\{|\Delta_k^n X| \leq \delta_{n,\epsilon}\}} \xrightarrow{\mathbb{P}} \int_0^1 g(X_t^0, \theta) dt, \quad (5.7)$$

*uniformly in  $\theta \in \Theta$ . In addition, suppose that Q2[M] holds for some  $M > 0$ , then*

$$\sup_{n \in \mathbb{N}, \epsilon > 0} \mathbb{R} \left[ \left( \epsilon^{-1} \sup_{\theta \in \Theta} \left| \frac{1}{n} \sum_{k=1}^n g_{k-1}(\theta) \mathbf{1}_{\{|\Delta_k^n X| \leq \delta_{n,\epsilon}\}} - \int_0^1 g(X_t^0, \theta) dt \right| \right)^M \right] < \infty. \quad (5.8)$$

*Proof.* Since  $|g(x)| \lesssim (1 + |x|)^C$ , we have that

$$\begin{aligned} & \sup_{\theta \in \Theta} \left| \frac{1}{n} \sum_{k=1}^n g_{k-1}(\theta) \mathbf{1}_{\{|\Delta_k^n X| \leq \delta_{n,\epsilon}\}} - \int_0^1 g(X_t^0, \theta) dt \right| \\ & \lesssim \sup_{\theta \in \Theta} \left| \frac{1}{n} \sum_{k=1}^n g_{k-1}(\theta) - \int_0^1 g(X_t^0, \theta) dt \right| + \frac{1}{n} \sum_{k=1}^n (1 + |X_{t_{k-1}^n}|)^C \mathbf{1}_{\{|\Delta_k^n X| > \delta_{n,\epsilon}\}} \\ & = \sup_{\theta \in \Theta} \left| \int_0^1 \{g(Y_t^{n,\epsilon}) - g(X_t^0, \theta)\} dt \right| + \frac{1}{n} \sum_{k=1}^n (1 + |X_{t_{k-1}^n}|)^C \mathbf{1}_{\{|\Delta_k^n X| > \delta_{n,\epsilon}\}}. \end{aligned}$$

On the last first term, it follows that

$$\begin{aligned} & \sup_{\theta \in \Theta} \left| \int_0^1 \{g(Y_t^{n,\epsilon}) - g(X_t^0, \theta)\} dt \right| \\ & \leq \sup_{\theta \in \Theta} \int_0^1 \int_0^1 |\nabla_x g(X_s^0 + u(Y_s^{n,\epsilon} - X_s^0), \theta)| \cdot |Y_s^{n,\epsilon} - X_s^0| du ds \\ & \lesssim \int_0^1 (1 + |X_s^0| + |Y_s^{n,\epsilon}|^C) |Y_s^{n,\epsilon} - X_s^0| ds \\ & \lesssim \left( 1 + \sup_{t \in [0,1]} |X_t^0| + \sup_{t \in [0,1]} |X_t^0| \right)^C \sup_{t \in [0,1]} |Y_t^{n,\epsilon} - X_t^0| \xrightarrow{\mathbb{P}} 0, \end{aligned} \quad (5.9)$$

by Lemma 1, (5.1). Hence the proof ends if we show the second term tends to zero in probability. Let

$$\xi_{k,n,\epsilon}(\theta) := \frac{1}{n} (1 + |X_{t_{k-1}^n}|)^C \mathbf{1}_{\{|\Delta_k^n X| > \delta_{n,\epsilon}\}}.$$

We show that

$$\sum_{k=1}^n \mathbb{E} \left[ \xi_{k,n,\epsilon}(\theta) | \mathcal{F}_{t_{k-1}^n} \right] \xrightarrow{\mathbb{P}} 0; \quad (5.11)$$

$$\sum_{k=1}^n \mathbb{E} \left[ |\xi_{k,n,\epsilon}(\theta)|^2 | \mathcal{F}_{t_{k-1}^n} \right] \xrightarrow{\mathbb{P}} 0. \quad (5.12)$$

which implies that  $\sum_{k=1}^n \xi_{k,n,\epsilon}(\theta) \xrightarrow{\mathbb{P}} 0$  for each  $\theta \in \Theta$  from Lemma 9 by Genon-Catalot and Jacod [7]. First, we show that (5.11). Note that

$$\sum_{k=1}^n \mathbb{E} \left[ \xi_{k,n,\epsilon}(\theta) | \mathcal{F}_{t_{k-1}^n} \right] = \frac{1}{n} \sum_{k=1}^n (1 + |X_{t_{k-1}^n}|)^C \mathbb{P} \left( |\Delta_k^n X| > \delta_{n,\epsilon} | \mathcal{F}_{t_{k-1}^n} \right)$$

Since it follows that  $n^{-1} \sum_{k=1}^n (1 + |X_{t_{k-1}^n}|)^C(\theta) = O_p(1)$  by Lemma 3.3 in [20], it suffices to show that

$$\mathbb{P} \left( |\Delta_k^n X| > \delta_{n,\epsilon} | \mathcal{F}_{t_{k-1}^n} \right) = o_p(1),$$

for any  $k = 1, \dots, n$ . Note that it holds that

$$\sup_{t \in (t_{k-1}^n, t_k^n]} |X_t - X_{t_{k-1}^n}| \lesssim \Delta_n (1 + \|X\|_*) + \epsilon \sup_{t \in (t_{k-1}^n, t_k^n]} |Q_t^\epsilon|.$$

Hence it follows from Q1[ $\gamma$ ] and the assumption on  $\delta_{n,\epsilon}$  that, for  $n$  large enough,

$$\begin{aligned} \mathbb{P} \left( |\Delta_k^n X| > \delta_{n,\epsilon} | \mathcal{F}_{t_{k-1}^n} \right) &\leq \mathbb{P} \left( \sup_{t \in (t_{k-1}^n, t_k^n]} |Q_t - Q_{t_{k-1}^n}| > \frac{\delta_{n,\epsilon}}{2\epsilon} \middle| \mathcal{F}_{t_{k-1}^n} \right) \\ &\quad + \mathbb{P} \left( (1 + \|X\|_*) > \frac{\delta_{n,\epsilon}}{2\Delta_n} \middle| \mathcal{F}_{t_{k-1}^n} \right) \\ &\leq \mathbb{P} \left( \sup_{t \in (t_{k-1}^n, t_k^n]} |Q_t - Q_{t_{k-1}^n}| \gtrsim \Delta_n^\gamma | \mathcal{F}_{t_{k-1}^n} \right) \\ &\quad + \mathbb{P} \left( (1 + \|X\|_*) \gtrsim \Delta_n^{-1} \delta_{n,\epsilon} | \mathcal{F}_{t_{k-1}^n} \right) \\ &= o_p(1), \end{aligned} \quad (5.13)$$

since  $\|X\|_*$  is bounded in probability. This is the proof of (5.11). The proof of (5.12) is similar to above, which ends the proof of (5.7). The proof of (5.8) is easy from the estimates (5.10) and Corollary 2 since now it holds that  $n\epsilon \rightarrow \infty$  (see Remark 2), so we omit the details. Then the proof is completed.  $\square$

**Lemma 4.** Let  $g \in C_{\uparrow}^{1,1}(\mathbb{R}^d \times \Theta; \mathbb{R}^d)$ . Under A1, it holds that

$$\sum_{k=1}^n g_{k-1}(\theta) [\Delta_k^n Q^\epsilon] \xrightarrow{\mathbb{P}} \int_0^1 g(X_t^0, \theta) [dQ_t].$$

*Proof.* Since we are assuming the uniform convergence (1.2), we have that

$$\left| \sum_{k=1}^n g_{k-1}(\theta) [\Delta_k^n Q^\epsilon] - \int_0^1 g(X_t^0, \theta) [dQ_t] \right|$$

$$\begin{aligned}
&= \left| \int_0^1 g(Y_t^{n,\epsilon}, \theta) [dQ_t^\epsilon - dQ_t] \right| + \left| \int_0^1 \{g(Y_t^{n,\epsilon}, \theta) - g(X_t^0, \theta)\} [dQ_t] \right| \\
&\leq \left| \int_0^1 g(Y_t^{n,\epsilon}, \theta) [dQ_t^\epsilon - dQ_t] \right| + \int_0^1 |g(Y_t^{n,\epsilon}, \theta) - g(X_t^0, \theta)| [|dA_t|] \\
&\quad + \left| \int_0^1 \{g(Y_t^{n,\epsilon}, \theta) - g(X_t^0, \theta)\} [dM_t] \right| \\
&=: I_{n,\epsilon}^{(1)} + I_{n,\epsilon}^{(2)} + I_{n,\epsilon}^{(3)}.
\end{aligned}$$

As for  $I_{n,\epsilon}^{(1)}$ : note that for  $t \leq \tau_m^{n,\epsilon}$ ,  $|Y_t^{n,\epsilon}| \leq m$ , hence  $|g(Y_t^{n,\epsilon}) \mathbf{1}_{\{t \leq \tau_m^{n,\epsilon}\}}| < C_m$  for some constant  $C_m$ , which is independent of  $n$  and  $\epsilon$ . Therefore it follows for any  $\eta > 0$  that

$$\begin{aligned}
\mathbb{P}(|I_{n,\epsilon}^{(1)}| > \eta) &\leq \mathbb{P}(\tau_m^{n,\epsilon} < 1) + \mathbb{P}\left(\left|\int_0^1 g(Y_t^{n,\epsilon}, \theta) \mathbf{1}_{\{t \leq \tau_m^{n,\epsilon}\}} [dQ_t^\epsilon - dQ_t]\right| > \eta/2\right) \\
&\lesssim \mathbb{P}(\tau_m^{n,\epsilon} < 1) + \mathbb{P}\left(\sum_{k=1}^n C_m \mathbf{1}_{\{t_{k-1}^n \leq t \leq t_k^n\}} |Q_t^\epsilon - Q_t| \gtrsim \eta\right) \\
&\lesssim \mathbb{P}(\tau_m^{n,\epsilon} < 1) + \mathbb{P}\left(\sup_{t \in [0,1]} |Q_t^\epsilon - Q_t| \gtrsim \eta\right) \rightarrow 0.
\end{aligned}$$

As for  $I_{n,\epsilon}^{(2)}$ , it follows from Lemma 1, (5.1) that

$$I_{n,\epsilon}^{(2)} \lesssim (1 + \sup_{t \in [0,1]} |X_t^0| + \sup_{t \in [0,1]} |X_t|)^C \sup_{t \in [0,1]} |Y_t^{n,\epsilon} - X_t^0| \xrightarrow{\mathbb{P}} 0.$$

As for  $I_{n,\epsilon}^{(3)}$ , using Markov's, and the Burkholder-Davis-Gundy inequalities, we have for any  $\eta > 0$  that

$$\begin{aligned}
\mathbb{P}(|I_{n,\epsilon}^{(3)}| > \eta) &\leq \mathbb{P}(\tau_m^{n,\epsilon} < 1) + \mathbb{P}\left(\left|\int_0^1 \{g(Y_t^{n,\epsilon}, \theta) - g(X_t^0, \theta)\} \mathbf{1}_{\{t \leq \tau_m^{n,\epsilon}\}} [dM_t]\right| > \eta/2\right) \\
&\leq \mathbb{P}(\tau_m^{n,\epsilon} < 1) + 2\eta^{-1} \mathbb{E} \left| \int_0^1 \{g(Y_t^{n,\epsilon}, \theta) - g(X_t^0, \theta)\} \mathbf{1}_{\{t \leq \tau_m^{n,\epsilon}\}} [dM_t] \right| \\
&\lesssim \mathbb{P}(\tau_m^{n,\epsilon} < 1) + \eta^{-1} \mathbb{E} \left| \int_0^1 |g(Y_t^{n,\epsilon}, \theta) - g(X_t^0, \theta)|^2 \mathbf{1}_{\{t \leq \tau_m^{n,\epsilon}\}} d[M, M]_t \right|^{1/2}.
\end{aligned}$$

From the definition of  $\tau_m^{n,\epsilon}$ , the integrand in the last term is bounded in  $n, \epsilon$ . Hence, taking the limit  $n \rightarrow \infty$  and  $\epsilon \rightarrow 0$ , we see from the dominated convergence theorem that  $\mathbb{P}(|I_{n,\epsilon}^{(3)}| > \eta) \rightarrow 0$ , which completes the proof.  $\square$

**Lemma 5.** Let  $g \in C_{\uparrow}^{1,1}(\mathbb{R}^d \times \Theta; \mathbb{R}^d)$ . Assume A1, A2, Q1[ $\gamma$ ], and that

$$\delta_{n,\epsilon} \Delta_n^{-1} \rightarrow \infty, \quad \epsilon \Delta_n^\gamma \delta_{n,\epsilon}^{-1} = O(1).$$

Then we have

$$\sum_{k=1}^n g_{k-1}(\theta) [\chi_k(\theta_0)] \mathbf{1}_{\{|\Delta_k^n X| \leq \delta_{n,\epsilon}\}} \xrightarrow{\mathbb{P}} 0, \quad (5.14)$$

uniformly in  $\theta \in \Theta$ . In addition, assume  $Q2[M]$  for any  $M > p$ . Then

$$\mathbb{E} \left[ \left( \epsilon^{-1} \sup_{\theta \in \Theta} \left| \sum_{k=1}^n g_{k-1}(\theta) [\chi_k(\theta_0)] \mathbf{1}_{\{|\Delta_k^n X| \leq \delta_{n,\epsilon}\}} \right| \right)^M \right] < \infty. \quad (5.15)$$

*Proof.* The proof of (5.14) is similar to the one of Lemma 3.5 by Long et al. [20] with a slight extension to semimartingale version; see also Remark 4.3 in [20]. It is clear from their proof that the indicator  $\mathbf{1}_{\{|\Delta_k^n X| \leq \delta_{n,\epsilon}\}}$  is not essential to the proof. So it is omitted.

As for (5.15), we note that

$$\begin{aligned} & \epsilon^{-1} \sup_{\theta \in \Theta} \left| \sum_{k=1}^n g_{k-1}(\theta) [\chi_k(\theta_0)] \mathbf{1}_{\{|\Delta_k^n X| \leq \delta_{n,\epsilon}\}} \right| \\ &= \epsilon^{-1} \sup_{\theta \in \Theta} \left| \sum_{k=1}^n \int_0^1 g(Y_s^{n,\epsilon}, \theta) [b(X_s, \theta_0) - b(Y_s^{n,\epsilon}, \theta_0) + \epsilon \cdot dQ_t^\epsilon] ds \cdot \mathbf{1}_{\{|\Delta_k^n X| \leq \delta_{n,\epsilon}\}} \right| \\ &\leq \epsilon^{-1} \int_0^1 \sup_{\theta \in \Theta} |g(Y_s^{n,\epsilon}, \theta) [b(X_s, \theta_0) - b(Y_s^{n,\epsilon}, \theta_0)]| ds \\ &\quad + \sup_{\theta \in \Theta} \left| \int_0^1 g(Y_s^{n,\epsilon}, \theta) [dQ_t^\epsilon - dQ_t] \right| + \sup_{\theta \in \Theta} \left| \int_0^1 g(Y_s^{n,\epsilon}, \theta) [dQ_t] \right| \\ &=: I_{n,\epsilon}^{(1)} + I_{n,\epsilon}^{(2)} + I_{n,\epsilon}^{(3)}. \end{aligned}$$

By the assumption A1 and the condition for  $g$ , we have

$$\begin{aligned} I_{n,\epsilon}^{(1)} &\lesssim \epsilon^{-1} \int_0^1 (1 + |Y_s^{n,\epsilon}|)^\lambda |X_s - Y_s^{n,\epsilon}| ds \\ &\lesssim \left( 1 + \|Y^{n,\epsilon} - X^0\|_*^\lambda + \|X^0\|_*^\lambda \right) (\epsilon^{-1} \|X - X^0\| + \epsilon^{-1} \|X^0 - Y^{n,\epsilon}\|) \end{aligned}$$

Hence, under the assumption (see Remark 2), Corollary 2 yields that

$$\mathbb{E} |I_{n,\epsilon}^{(1)}|^M < \infty.$$

We have already shown that  $I_{n,\epsilon}^{(2)} \xrightarrow{\mathbb{P}} 0$  in the proof of Lemma 4. Hence the proof ends if we show that  $I_{n,\epsilon}^{(3)} \xrightarrow{\mathbb{P}} 0$ . Noticing that a bounded convex set  $\Theta$  admits the following *Sobolev inequality*;

$$\sup_{\theta \in \Theta} |u(\theta)| \lesssim \|u(\theta)\|_{L^q(\Theta)} + \|\nabla_\theta u(\theta)\|_{L^q(\Theta)},$$

for  $q > p = \dim(\Theta)$ , we see for any  $M > p$  that

$$\begin{aligned} I_{n,\epsilon}^{(3)} &\leq \int_0^1 (1 + |Y_s^{n,\epsilon}, \theta|)^\lambda \cdot |dA_s| + \sup_{\theta \in \Theta} \left| \int_0^1 g(Y_s^{n,\epsilon}, \theta) [dM] \right| \\ &\lesssim \left( 1 + \|Y^{n,\epsilon} - X^0\|_*^\lambda + \|X^0\|_*^\lambda \right) TV(A) + \left( \int_\Theta \left| \int_0^1 g(Y_s^{n,\epsilon}, \theta) [dM^\epsilon] \right|^N d\theta \right)^{1/M}. \end{aligned}$$

Then, by using the Burkholder-Davis-Gundy inequality, we have

$$\begin{aligned} \mathbb{E} |I_{n,\epsilon}^{(3)}|^M &\lesssim \mathbb{E} \left[ \left( 1 + \|Y^{n,\epsilon} - X^0\|_*^{\lambda M} \right) TV(A)^M \right] + \int_\Theta \mathbb{E} \left| \int_0^1 g(Y_s^{n,\epsilon}, \theta) [dM] \right|^M d\theta \\ &\lesssim 1 + \mathbb{E} \left[ \left( 1 + \|Y^{n,\epsilon} - X^0\|_*^\lambda + \|X^0\|_*^\lambda \right)^M |[M, M]_1|^{M/2} \right] < \infty, \end{aligned}$$

under Q2[M] for any  $M > p$ . This completes the proof of (5.15).  $\square$

**Lemma 6.** Let  $g \in C_{\uparrow}^{1,1}(\mathbb{R}^d \times \Theta; \mathbb{R}^d)$ . Assume A1, A2, Q1[ $\gamma$ ], and that

$$\delta_{n,\epsilon} \Delta_n^{-1} \rightarrow \infty, \quad \epsilon \Delta_n^\gamma \delta_{n,\epsilon}^{-1} = O(1)$$

Then we have

$$\sum_{k=1}^n g_{k-1}(\theta) [\chi_k(\theta_0)] \mathbf{1}_{\{|\Delta_k^n X| > \delta_{n,\epsilon}\}} = o_p(\epsilon),$$

for each  $\theta \in \Theta$ .

*Proof.* Using  $\Delta_k^n X = \int_{t_{k-1}^n}^{t_k^n} b(X_t, \theta_0) dt + \epsilon \Delta_k^n Q^\epsilon$ , we have that

$$\begin{aligned} &\epsilon^{-1} \sum_{k=1}^n g_{k-1}(\theta) [\chi_k(\theta_0)] \mathbf{1}_{\{|\Delta_k^n X| > \delta_{n,\epsilon}\}} \\ &= \sum_{k=1}^n \epsilon^{-1} \int_{t_{k-1}^n}^{t_k^n} g_{k-1}(\theta) [b(X_t, \theta_0) - b_{k-1}(\theta_0)] dt \cdot \mathbf{1}_{\{|\Delta_k^n X| > \delta_{n,\epsilon}\}} \\ &\quad + \sum_{k=1}^n g_{k-1}(\theta) [\Delta_k^n Q^\epsilon] \cdot \mathbf{1}_{\{|\Delta_k^n X| > \delta_{n,\epsilon}\}} \\ &=: H_{n,\epsilon}^{(1)}(\theta) + H_{n,\epsilon}^{(2)}(\theta). \end{aligned}$$

As for  $H_{n,\epsilon}^{(1)}$ , we easily find it converges to zero in probability as  $n\epsilon \rightarrow \infty$  (see Remark 2) by the same argument as for  $H_{n,\epsilon}^{(1)}$  in Lemma 3.6 by Long et al. [20]. As for  $H_{n,\epsilon}^{(2)}$ : since

$$\sum_{k=1}^n g_{k-1}(\theta) [\Delta_k^n Q^\epsilon] \xrightarrow{\mathbb{P}} \int_0^1 g(X_t^0, \theta) [dQ_t],$$

for each  $\theta \in \Theta$  by Lemma 4,  $H_{n,\epsilon}^{(2)}(\theta) \xrightarrow{\mathbb{P}} 0$  for each  $\theta \in \Theta$  if

$$\mathbb{P}(|\Delta_k^n X| > \delta_{n,\epsilon}) \rightarrow 0 \quad \text{for any } k,$$

which is true by (5.13). Hence the proof is completed.  $\square$

## 5.2 Proof of Theorem 1

We shall show that  $\hat{\theta}_{n,\epsilon}$  is asymptotically equivalent to  $\hat{\theta}_{n,\epsilon}^{LSE}$  given in (1.4). Let

$$\begin{aligned}\tilde{\Psi}_{n,\epsilon}(\theta) &:= \epsilon^2 (\Psi_{n,\epsilon}(\theta) - \Psi_{n,\epsilon}(\theta_0)), \\ \tilde{\Phi}_{n,\epsilon}(\theta) &:= \epsilon^2 (\Phi_{n,\epsilon}(\theta) - \Phi_{n,\epsilon}(\theta_0)).\end{aligned}$$

where  $\Psi_{n,\epsilon}$  and  $\Phi_{n,\epsilon}$  are given in (1.5) and (1.7). Then  $\hat{\theta}_{n,\epsilon}^{LSE}$  and  $\hat{\theta}_{n,\epsilon}$  are respectively minimum contrast estimators for contrast functions  $\Psi_{n,\epsilon}$  and  $\Phi_{n,\epsilon}$ .

By the same argument as in the proof of Theorem 2.1 with Remark 4.3 by Long et al. [20], all we need to show is

$$\sup_{\theta \in \Theta} |\tilde{\Phi}_{n,\epsilon}(\theta) - F(\theta_0)| \xrightarrow{\mathbb{P}} 0,$$

where  $F(\theta) := \int_0^1 |b(X_t^0, \theta) - b(X_t^0, \theta_0)|^2 dt$ . Since  $\sup_{\theta \in \Theta} |\tilde{\Psi}_{n,\epsilon}(\theta) - F(\theta_0)| \xrightarrow{\mathbb{P}} 0$ , Therefore,

$$\begin{aligned}\sup_{\theta \in \Theta} |\tilde{\Phi}_{n,\epsilon}(\theta) - F(\theta_0)| &\leq \sup_{\theta \in \Theta} |\tilde{\Phi}_{n,\epsilon}(\theta) - \tilde{\Psi}_{n,\epsilon}(\theta)| + \sup_{\theta \in \Theta} |\tilde{\Psi}_{n,\epsilon}(\theta) - F(\theta_0)| \\ &= \sup_{\theta \in \Theta} \left| n \sum_{k=1}^n |\chi_k(\theta)|^2 \mathbf{1}_{\{|\Delta_k^n X| > \delta_{n,\epsilon}\}} \right| + o_p(1) \\ &\lesssim \sup_{\theta \in \Theta} \left| \sum_{k=1}^n (b_{k-1}(\theta) - b_{k-1}(\theta_0)) [\chi_k(\theta_0)] \mathbf{1}_{\{|\Delta_k^n X| > \delta_{n,\epsilon}\}} \right| \\ &\quad + \sup_{\theta \in \Theta} \left| \frac{1}{n} \sum_{k=1}^n |b_{k-1}(\theta) - b_{k-1}(\theta_0)|^2 \mathbf{1}_{\{|\Delta_k^n X| > \delta_{n,\epsilon}\}} \right| + o_p(1)\end{aligned}$$

The last first and second terms converges to zero in probability by Lemmas 3 and 5. This completes the proof.

## 5.3 Proof of Theorem 2

We use the following notation:

- $G_{n,\epsilon}(\theta) = 2^{-1} \nabla_{\theta} \Phi_{n,\epsilon}(\theta) \in \mathbb{R}^p$ ;
- $K_{n,\epsilon}(\theta) = \nabla_{\theta} G_{n,\epsilon}(\theta) (= \nabla_{\theta}^2 \Phi_{n,\epsilon}(\theta)) \in \mathbb{R}^p \otimes \mathbb{R}^p$ ;

- $K(\theta) = \int_0^1 \nabla_\theta^2 b(X_t^0, \theta) [b(X_t^0, \theta_0) - b(X_t^0, \theta)] dt - I(\theta_0) \in \mathbb{R}^p \otimes \mathbb{R}^p.$

Then it follows by Taylor's formula that, for some  $\rho \in (0, 1)$ ,

$$\int_0^1 K_{n,\epsilon} \left( \theta_0 + u(\hat{\theta}_{n,\epsilon} - \theta_0) \right) du \cdot \epsilon^{-1} (\hat{\theta}_{n,\epsilon} - \theta_0) = \epsilon^{-1} G_{n,\epsilon}(\hat{\theta}_{n,\epsilon}) - \epsilon^{-1} G_{n,\epsilon}(\theta_0).$$

Let us show that

$$\epsilon^{-1} G_{n,\epsilon}(\theta_0) \xrightarrow{\mathbb{P}} \int_0^1 \nabla_\theta b(X_t^0, \theta_0) [dQ_t]; \quad (5.16)$$

$$\sup_{\theta \in \Theta} |K_{n,\epsilon}(\theta) - K(\theta)| \xrightarrow{\mathbb{P}} 0. \quad (5.17)$$

Then the result follows by the same argument as in the proof of Theorem 2.2 by Long et al. [20]; see also Uchida [34]. As for (5.16): it follows that

$$\begin{aligned} \epsilon^{-1} G_{n,\epsilon}(\theta_0) &= \epsilon^{-1} \sum_{k=1}^n \nabla_\theta b_{k-1}(\theta_0) [\chi_k(\theta_0)] \mathbf{1}_{\{|\Delta_k^n X| \leq \delta_{n,\epsilon}\}} \\ &= \epsilon^{-1} \sum_{k=1}^n \nabla_\theta b_{k-1}(\theta_0) [\chi_k(\theta_0)] - \epsilon^{-1} \sum_{k=1}^n \nabla_\theta b_{k-1}(\theta_0) [\chi_k(\theta_0)] \mathbf{1}_{\{|\Delta_k^n X| > \delta_{n,\epsilon}\}} \\ &= \epsilon^{-1} \sum_{k=1}^n \nabla_\theta b_{k-1}(\theta_0) \left[ \int_{t_{k-1}^n}^{t_k^n} \{b(X_t, \theta_0) - b_{k-1}(\theta_0)\} dt \right] + \sum_{k=1}^n \nabla_\theta b_{k-1}(\theta_0) [\Delta_k^n Q] \\ &\quad - \epsilon^{-1} \sum_{k=1}^n \nabla_\theta b_{k-1}(\theta_0) [\chi_k(\theta_0)] \mathbf{1}_{\{|\Delta_k^n X| > \delta_{n,\epsilon}\}} \end{aligned}$$

Then we can show that the last first term converges to zero in probability by the same evaluation as for  $H_{n,\epsilon}^{(1)}(\theta_0)$  in the proof of Lemma 3.6 in [20], the second term converges to  $\int_0^1 \nabla_\theta b(X_t^0, \theta_0) [dQ_t]$  in probability by Lemma 4, and that the third term goes to zero in probability by Lemma 6. Similarly, as for (5.17), it follows that

$$\begin{aligned} K_{n,\epsilon}(\theta) &= \sum_{k=1}^n \nabla_\theta^2 b_{k-1}(\theta) [\chi_k(\theta)] \mathbf{1}_{\{|\Delta_k^n X| \leq \delta_{n,\epsilon}\}} - \frac{1}{n} \sum_{k=1}^n \nabla_\theta b_{k-1}(\theta)^\top \nabla_\theta b_{k-1}(\theta) \mathbf{1}_{\{|\Delta_k^n X| \leq \delta_{n,\epsilon}\}} \\ &= \sum_{k=1}^n \nabla_\theta^2 b_{k-1}(\theta) [\chi_k(\theta)] \mathbf{1}_{\{|\Delta_k^n X| \leq \delta_{n,\epsilon}\}} \\ &\quad + \frac{1}{n} \sum_{k=1}^n \nabla_\theta^2 b_{k-1}(\theta) [b_{k-1}(\theta_0) - b_{k-1}(\theta)] \mathbf{1}_{\{|\Delta_k^n X| \leq \delta_{n,\epsilon}\}} \\ &\quad - \frac{1}{n} \sum_{k=1}^n \nabla_\theta b_{k-1}(\theta)^\top \nabla_\theta b_{k-1}(\theta) \mathbf{1}_{\{|\Delta_k^n X| \leq \delta_{n,\epsilon}\}}, \end{aligned}$$

cf. the expression of  $K_{n,\epsilon}^{ij}(\theta)$  in the proof of Lemma 3.7 by Long et al. [20]. Hence Lemmas 3 and 5 yield (5.17). Using the facts (5.16) and (5.17), and the consistency result: Theorem 1, we can show that

$$\epsilon^{-1}(\widehat{\theta}_{n,\epsilon} - \theta_0) \sim^p -K^{-1}(\theta_0) \cdot \epsilon^{-1}G_{n,\epsilon}(\theta_0), \quad n \rightarrow \infty, \quad \epsilon \rightarrow 0,$$

by completely the same argument as in the proof of Theorem 2.2 by Long et al. [20]. Therefore, the proof is completed.

#### 5.4 Proof of Theorem 3

Denote by  $\widehat{u} := \epsilon^{-1}(\widehat{\theta}_{n,\epsilon} - \theta_0)$ . Since  $\widehat{u} \xrightarrow{\mathbb{P}} \zeta$  from Theorem 2, it suffices for the consequence to show that  $\widehat{u}$  is  $L^p$ -bounded:  $\sup_{n,\epsilon} \mathbb{E}|\widehat{u}|^p < \infty$  for any  $p > 0$ . For this proof, let  $U_{n,\epsilon}(\theta_0) = \{u \in \mathbb{R}^p : \theta_0 + \epsilon u \in \Theta_0\}$ , and define random fields  $\mathbb{Z}_{n,\epsilon} : U_{n,\epsilon}(\theta_0) \rightarrow \mathbb{R}_+$  by

$$\mathbb{Z}_{n,\epsilon}(u) = \exp\{-\Phi_{n,\epsilon}(\theta_0 + \epsilon u) + \Phi_{n,\epsilon}(\theta_0)\}, \quad u \in U_{n,\epsilon}(\theta_0).$$

Then, since  $\theta_0 \in \Theta_0$ , we see that

$$\mathbb{Z}_{n,\epsilon}(\widehat{u}) \geq \sup_{u \in U_{n,\epsilon}(\theta_0)} \mathbb{Z}_{n,\epsilon}(u) \geq \mathbb{Z}_{n,\epsilon}(0) = 1. \quad (5.18)$$

Setting  $V_{n,\epsilon}(r) := U_{n,\epsilon}(\theta_0) \cap \{u \in \mathbb{R}^p : |u| \geq r\}$ , we consider the following condition for the random fields  $\mathbb{Z}_{n,\epsilon}$ : for every  $L > 0$  and  $r > 0$ ,

$$\mathbb{P}\left(\sup_{u \in V_{n,\epsilon}(r)} \mathbb{Z}_{n,\epsilon}(u) \geq e^{-r}\right) \lesssim r^{-L}, \quad (5.19)$$

which is called the *polynomial type large deviation inequality (PLDI)*, and is investigated by Yoshida [38] in details. If this PLDI holds true then, for any  $L > p$ ,

$$\begin{aligned} \sup_{n \in \mathbb{N}, \epsilon > 0} \mathbb{E}|\widehat{u}|^p &= \sup_{n \in \mathbb{N}, \epsilon > 0} p \int_0^\infty r^{p-1} \mathbb{P}(|\widehat{u}| \geq r) \, dr \\ &\leq \sup_{n \in \mathbb{N}, \epsilon > 0} p \int_0^\infty r^{p-1} \left\{ 1 \wedge \mathbb{P}\left(\sup_{u \in V_{n,\epsilon}(r)} \mathbb{Z}_{n,\epsilon}(u) \geq 1\right) \right\} \, dr \\ &\lesssim \int_0^\infty r^{p-1} (1 \wedge r^{-L}) \, dr < \infty, \end{aligned}$$

here we used (5.18) in the first inequality. Therefore the proof ends if we show (5.19), some sufficient conditions for which are found in the paper by Yoshida [38]. Here we shall verify the conditions [A1''], [A4'], [A6], [B1] and [B2] given in Theorem 3, (c) in [38]. See also Ogihara and Yoshida [21] or Masuda [19] for simplified descriptions for those conditions.

Applying Taylor's formula with the notation  $G_{n,\epsilon}(\theta)$ ,  $K_{n,\epsilon}(\theta)$  and  $K(\theta)$  given in the proof of Theorem 2, we have

$$\log \mathbb{Z}_{n,\epsilon}(u) = -\Phi_{n,\epsilon}(\theta_0 + \epsilon u) + \Phi_{n,\epsilon}(\theta_0)$$



$$= -\epsilon G_{n,\epsilon}(\theta_0)[u] - \frac{\epsilon^2}{2} \{-K(\theta_0)\} [u^{\otimes 2}] + R_{n,\epsilon}(u),$$

where

$$\begin{aligned} R_{n,\epsilon}(u) &= \epsilon^2 \int_0^1 (s-1) \{K(\theta_0)[u^{\otimes 2}] - K_{n,\epsilon}(\theta_0 + s \cdot \epsilon u)[u^{\otimes 2}]\} ds \\ &= \frac{\epsilon^2}{2} \{K_{n,\epsilon}(\theta_0) - K(\theta_0)\} [u^{\otimes 2}] - \epsilon^3 \int_0^1 (s-1) \int_0^1 \nabla_\theta K_{n,\epsilon}(\theta_0 + ts \cdot \epsilon u)[u^{\otimes 3}] dt ds. \end{aligned}$$

This means that  $\mathbb{Z}_{n,\epsilon}$  could be *Partially Locally Asymptotically Quadratic (PLAQ)*, which is a starting point of [38]. According to Theorem 3, (c) in [38], if we take some “tuning parameters” given in [A4'] in [38] such as  $\beta_1 \approx 1/2$ ,  $\rho_1, \rho_2, \beta, \beta_2 \approx 0$ , then the PLDI (5.19) holds true if the following [A1''], [A6], [B1] and [B2] are satisfied; we use the same conditioning numbers as in [38] to make those correspondences clear.

[A1''] For every  $M > 0$ ,

$$\sup_{n \in \mathbb{N}, \epsilon > 0} \mathbb{E} \left[ \left( \epsilon^2 \sup_{\theta \in \Theta} |\nabla_\theta^3 \Phi_{n,\epsilon}(\theta)| \right)^M \right] < \infty. \quad (5.20)$$

Moreover, for given  $L > 0$  and any  $\delta > 0$  small enough,

$$\sup_{n \in \mathbb{N}, \epsilon > 0} \mathbb{E} \left[ \left( \epsilon^{-1} |K_{n,\epsilon}(\theta_0) - K(\theta_0)| \right)^{L-\delta} \right] < \infty. \quad (5.21)$$

[A6] For any  $\delta > 0$  small enough,

$$\sup_{n \in \mathbb{N}, \epsilon > 0} \mathbb{E} \left[ |\epsilon G_{n,\epsilon}(\theta_0)|^{L+\delta} \right] < \infty; \quad (5.22)$$

$$\sup_{n \in \mathbb{N}, \epsilon > 0} \mathbb{E} \left[ \sup_{\theta \in \Theta} \left( \epsilon^{-1} |\tilde{\Phi}_{n,\epsilon}(\theta) - F(\theta)| \right)^{L+\delta} \right] < \infty, \quad (5.23)$$

where  $\tilde{\Phi}_{n,\epsilon}$  and  $F(\theta) = \int_0^1 |b(X_t^0, \theta) - b(X_t^0, \theta_0)|^2 dt$  are given in the proof of Theorem 1.

[B1] The matrix  $-K(\theta_0)$  ( $= I(\theta_0)$ ) is deterministic and positive definite.

[B2] There exists a deterministic positive number  $\chi$  such that

$$-\{F(\theta) - F(\theta_0)\} \leq -\chi |\theta - \theta_0|^2.$$

(Note that the notational correspondence between [38] and ours is:  $a_T = \epsilon$ ;  $b_T = \epsilon^{-2}$ ;  $\mathbb{H}_T = -\Phi_{n,\epsilon}$ ;  $\mathbb{Y} = -F$  and  $\Gamma = -K$ ).

Now we can easily check that the conditions (5.20) and (5.22) are true by Lemma 5, (5.15), and that (5.21) and (5.23) are also true by Lemma 3, (5.8). Moreover the conditions [B1] and [B2] are clear from the assumptions A1 and A4, respectively, the proof ends if we show (5.20)–(5.23). Hence the proof is completed.

### 5.5 Proof of Theorem 4

First, we shall show the consistency:

$$\tilde{\theta}_{n,\epsilon,\delta} \xrightarrow{\mathbb{P}} \theta_0.$$

Since we suppose that the jumps are specified, the following “negligibility” is obtained:

$$\begin{aligned} \mathbb{P}(\|\Delta Q^\epsilon\|_k^* > \delta) &= 1 - \mathbb{P}\left\{\sup_{t \in (t_{k-1}^n, t_k^n]} |\Delta Q_t| \leq \delta/\epsilon\right\} \\ &= 1 - e^{-\lambda(\delta/\epsilon)\Delta_n} \rightarrow 0 \end{aligned} \quad (5.24)$$

since  $\lambda(\delta/\epsilon)\Delta_n \rightarrow 0$ . Then we have the following lemma, which is the same type of results as Lemmas 3 and 5.

**Lemma 7.** *Let  $g \in C_{\uparrow}^{1,1}(\mathbb{R}^d \times \Theta; \mathbb{R})$ . Suppose A1, A2, and that*

$$\delta/\epsilon \rightarrow 0.$$

*Then, we have*

$$\begin{aligned} \frac{1}{n} \sum_{k=1}^n g_{k-1}(\theta) \mathbf{1}_{\{\|\Delta Q^\epsilon\|_k^* \leq \delta\}} &\xrightarrow{\mathbb{P}} \int_0^1 g(X_t^0, \theta) dt, \\ \sum_{k=1}^n g_{k-1}(\theta) [\chi_k(\theta_0)] \mathbf{1}_{\{\|\Delta Q^\epsilon\|_k^* \leq \delta\}} &\xrightarrow{\mathbb{P}} 0, \end{aligned}$$

*uniformly in  $\theta \in \Theta$ .*

The proof of this lemma is an obvious modification of the proofs of Lemmas 3 and 5 using the negligibility condition (5.24). Then by the same argument as in the proof of Theorem 1, the consistency follows.

Next, note that, as in the proof of Theorem 2,

$$\int_0^1 K_{n,\epsilon,\delta}(\theta_0 + u(\tilde{\theta}_{n,\epsilon,\delta} - \theta_0)) du \cdot (\tilde{\theta}_{n,\epsilon,\delta} - \theta_0) = G_{n,\epsilon,\delta}(\tilde{\theta}_{n,\epsilon,\delta}) - G_{n,\epsilon,\delta}(\theta_0).$$

where

$$\begin{aligned} G_{n,\epsilon,\delta}(\theta) &= 2^{-1} \nabla_{\theta} \tilde{\Phi}_{n,\epsilon,\delta}(\theta); \quad K_{n,\epsilon,\delta}(\theta) = \nabla_{\theta} G_{n,\epsilon,\delta}(\theta); \\ K(\theta) &= \int_0^1 \nabla_{\theta}^2 b(X_t^0, \theta) [b(X_t^0, \theta_0) - b(X_t^0, \theta)] dt - I(\theta_0). \end{aligned}$$

If we show that

$$\epsilon^{-1} G_{n,\epsilon,\delta}(\theta_0) \xrightarrow{\mathbb{P}} \int_0^1 \nabla_{\theta} b(X_t^0, \theta_0) [dW_t], \quad (5.25)$$

then we obtain the consequence because the convergence

$$\sup_{\theta \in \Theta} |K_{n,\epsilon,\delta}(\theta) - K(\theta)| \xrightarrow{\mathbb{P}} 0$$

holds true due to Lemma 7 and the same argument as in the proof of Lemma 3.7 in [20].

Note that

$$\begin{aligned} \epsilon^{-1} G_{n,\epsilon,\delta}(\theta_0) &= \epsilon^{-1} \sum_{k=1}^n \nabla_{\theta} b_{k-1}(\theta_0) \left[ \int_{t_{k-1}^n}^{t_k^n} \{b(X_s, \theta_0) - b_{k-1}(\theta_0)\} dt \right] \mathbf{1}_{\{\|\Delta Q^{\epsilon}\|_k^* \leq \delta\}} \\ &\quad + \sum_{k=1}^n \nabla_{\theta} b_{k-1}(\theta_0) \left[ \sum_{i=N_{t_{k-1}^n}^n+1}^{N_{t_k^n}^n} Y_i \mathbf{1}_{\{|Y_i| \leq \delta\}} \right] \\ &\quad + \sum_{k=1}^n \nabla_{\theta} b_{k-1}(\theta_0) [\Delta_k^n W] \mathbf{1}_{\{\|\Delta Q^{\epsilon}\|_k^* \leq \delta\}} \\ &=: I_{n,\epsilon,\delta}^{(1)} + I_{n,\epsilon,\delta}^{(2)} + I_{n,\epsilon,\delta}^{(3)}. \end{aligned}$$

Note that  $I_{n,\epsilon,\delta}^{(1)} \rightarrow 0$  by the same argument as in the proof of Lemma 3.6,  $H_{n,\epsilon}^{(1)}(\theta_0)$  in [20] by the assumption (4.5). Moreover, it is easy to see that

$$\mathbb{E} \left| I_{n,\epsilon,\delta}^{(2)} \right| = \frac{\lambda \delta}{n \epsilon} \sum_{k=1}^n \mathbb{E} |\nabla_{\theta} b_{k-1}(\theta_0)| \rightarrow 0.$$

Furthermore, note that

$$I_{n,\epsilon,\delta}^{(3)} = \sum_{k=1}^n \nabla_{\theta} b_{k-1}(\theta_0) [\Delta_k^n W] - \sum_{k=1}^n \nabla_{\theta} b_{k-1}(\theta_0) [\Delta_k^n W] \mathbf{1}_{\{\|\Delta Q^{\epsilon}\|_k^* > \delta\}}$$

Now, observe a measurability that

$$\{\omega \in \Omega : \|\Delta Q\|_k^* \leq \delta\} \in \sigma\left(N((t_{k-1}^n, s], [-\delta, \delta]) : s \in (t_{k-1}^n, t_k^n]\right),$$

which is independent of  $\Delta_k^n W$  and  $X_{t_{k-1}^n}^n$ . Therefore,

$$\begin{aligned} &\mathbb{E} \left| \sum_{k=1}^n \nabla_{\theta} b_{k-1}(\theta_0) [\Delta_k^n W] \mathbf{1}_{\{\|\Delta Q^{\epsilon}\|_k^* > \delta\}} \right| \\ &\leq \sum_{k=1}^n \mathbb{E} |\nabla_{\theta} b_{k-1}(\theta_0) [\Delta_k^n W]| \mathbb{P}(\|\Delta Q^{\epsilon}\|_k^* > \delta) \\ &\leq \sum_{k=1}^n \left( \mathbb{E} \int_{t_{k-1}^n}^{t_k^n} \text{trace}(\nabla_{\theta} b_{k-1}^{\otimes 2}(\theta_0)) dt \right)^{1/2} \left( 1 - e^{-\lambda(\delta/\epsilon)\Delta_n} \right) \rightarrow 0 \end{aligned}$$

from (5.24). As a result, Lemma (4) yields that

$$I_{n,\epsilon,\delta}^{(3)} \xrightarrow{\mathbb{P}} \int_0^1 \nabla_{\theta} b(X_t^0, \theta_0) [dW_t].$$

This completes the proof of (5.25).

## 5.6 Proof of Theorem 5

Note that, under the asymptotic conditions, it follows that

$$\mathbb{P}(\|\Delta Q^{\epsilon}\|_k^* \leq \delta) = e^{-\lambda(\delta/\epsilon)\Delta_n} \rightarrow 0; \quad \mathbb{P}(\|\Delta Q^{\epsilon}\|_k^* > \delta) \rightarrow 1, \quad (5.26)$$

which is a different situation in the previous theorem.

Under this setting, we can show the following lemma.

**Lemma 8.** *Let  $g \in C_{\uparrow}^{1,1}(\mathbb{R}^d \times \Theta; \mathbb{R})$ . Suppose A1, A2, Q1[ $\gamma$ ], and that*

$$\frac{\lambda(\delta/\epsilon)}{n \log n} \rightarrow c \in (0, 1), \quad n\epsilon\Delta_n^{\gamma} \rightarrow 0.$$

*Then it follows that, for  $\eta_n := e^{-\lambda(\delta/\epsilon)\Delta_n}$ ,*

$$\frac{1}{n\eta_n} \sum_{k=1}^n g_{k-1}(\theta) \mathbf{1}_{\{\|\Delta Q^{\epsilon}\|_k^* \leq \delta\}} \xrightarrow{\mathbb{P}} \int_0^1 g(X_t^0, \theta) dt, \quad (5.27)$$

$$\frac{1}{\eta_n} \sum_{k=1}^n g_{k-1}(\theta) [\chi_k(\theta_0)] \mathbf{1}_{\{\|\Delta Q^{\epsilon}\|_k^* \leq \delta\}} \xrightarrow{\mathbb{P}} 0, \quad (5.28)$$

*uniformly in  $\theta \in \Theta$ .*

*Proof of Lemma 8.*

As for (5.27): note that

$$\begin{aligned} & \sup_{\theta \in \Theta} \left| \frac{1}{n\eta_n} \sum_{k=1}^n g_{k-1}(\theta) \mathbf{1}_{\{\|\Delta Q^{\epsilon}\|_k^* \leq \delta\}} \xrightarrow{\mathbb{P}} \int_0^1 g(X_t^0, \theta) dt \right| \\ & \leq \sup_{\theta \in \Theta} \left| \frac{1}{n\eta_n} \sum_{k=1}^n g_{k-1}(\theta) \mathbf{1}_{\{\|\Delta Q^{\epsilon}\|_k^* \leq \delta\}} - \frac{1}{n} \sum_{k=1}^n g_{k-1}(\theta) \right| \\ & \quad + \sup_{\theta \in \Theta} \left| \frac{1}{n} \sum_{k=1}^n g_{k-1}(\theta) - \int_0^1 g(X_t^0, \theta) dt \right|. \end{aligned}$$

and the last second term converges to zero in probability from Lemma 3.3 by Long et al. [20]. The last first term is rewritten as  $\sum_{k=1}^n \xi_k^n(\theta)$  with

$$\xi_k^n(\theta) = \frac{1}{n} g_{k-1}(\theta) \left( \frac{1}{\eta_n} \mathbf{1}_{\{\|\Delta Q^{\epsilon}\|_k^* \leq \delta\}} - 1 \right).$$

Then we immediately see that, since  $\mathcal{F}_{t_{k-1}^n}$  and  $\|\Delta Q^\epsilon\|_k$  are independent each other,

$$\sum_{k=1}^n \mathbb{E}[\xi_k^n(\theta) | \mathcal{F}_{t_{k-1}^n}] = \frac{1}{n} \sum_{k=1}^n g_{k-1}(\theta) [\eta_n^{-1} \mathbb{P}(\|\Delta Q^\epsilon\|_k^* > \delta) - 1] = 0.$$

Moreover

$$\begin{aligned} \sum_{k=1}^n \mathbb{E}[|\xi_k^n(\theta)|^2 | \mathcal{F}_{t_{k-1}^n}] &= \frac{1}{n^2} \sum_{k=1}^n g_{k-1}^2(\theta) \mathbb{E} \left[ \left| \frac{1}{\eta_n} \mathbf{1}_{\{\|\Delta Q^\epsilon\|_k^* \leq \delta\}} - 1 \right|^2 \right] \\ &= O_p \left( \frac{1}{n\eta_n} \right) = O_p \left( \frac{1}{e^{(1-c_2)\log n}} \right) \xrightarrow{\mathbb{P}} 0, \end{aligned}$$

by the assumption. Hence we have that  $\sum_{k=1}^n \xi_k^n(\theta) \xrightarrow{\mathbb{P}} 0$  for every  $\theta \in \Theta$  from Lemma 9 by Genon-Catalot and Jacod [7].

To prove the uniformity of convergence, we have to show the tightness of the sequence  $\{\sum_{k=1}^n \xi_k^n(\cdot)\}_n$ . We shall use Theorem 20 in Appendix 1 by Ibragimov and Has'minskii [9], that is, we shall show that, for some  $H > 0$  and any  $N \in \mathbb{N}$ ,

$$\mathbb{E} \left| \sum_{k=1}^n \xi_k^n(\theta) \right|^{2N} < H; \quad (5.29)$$

$$\mathbb{E} \left| \sum_{k=1}^n [\xi_k^n(\theta_1) - \xi_k^n(\theta_2)] \right|^{2N} \leq H |\theta_1 - \theta_2|^{2N} \quad (5.30)$$

As for (5.29): using the independent property of  $X_{t_{k-1}^n}$  and  $\|\Delta Q^\epsilon\|_k$ , we have that

$$\begin{aligned} \mathbb{E} \left| \sum_{k=1}^n \xi_k^n(\theta) \right|^{2N} &= \sum_{k=1}^n \mathbb{E} \left[ |g_{k-1}(\theta)|^{2N} \mathbb{E} \left[ \left| \frac{1}{\eta_n} \mathbf{1}_{\{\|\Delta Q^\epsilon\|_k^* \leq \delta\}} - 1 \right|^{2N} | \mathcal{F}_{t_{k-1}^n} \right] \right] \\ &\lesssim \mathbb{E} \left[ \sup_{\theta \in \Theta, t \in [0,1]} |g(X_t)|^{2N} \right] \frac{1}{n^{2N}} \sum_{k=1}^n \left( \frac{1}{\eta_n^{2N-1}} + 1 \right) \\ &= O \left( \frac{1}{(n\eta_n)^{2N-1}} \right) = O \left( n^{-(1-c)} \right) \rightarrow 0. \end{aligned}$$

Therefore it is bounded. Inequality (5.30) is similarly proved since  $g \in C_{\uparrow}^{1,1}(\mathbb{R}^d \times \Theta; \mathbb{R})$ . Hence (5.27) is proved.

Finally we shall show (5.28). Note that

$$\begin{aligned} &\sup_{\theta \in \Theta} \left| \frac{1}{\eta_n} \sum_{k=1}^n g_{k-1}(\theta) [\chi_k(\theta_0)] \mathbf{1}_{\{\|\Delta Q^\epsilon\|_k^* \leq \delta\}} \right| \\ &\leq \frac{1}{\eta_n} \sum_{k=1}^n \int_{t_{k-1}^n}^{t_k^n} \sup_{\theta \in \Theta} |g_{k-1}(\theta) [b(X_s, \theta_0) - b_{k-1}(\theta_0)]| \, ds \cdot \mathbf{1}_{\{\|\Delta Q^\epsilon\|_k^* \leq \delta\}} \end{aligned}$$

$$\begin{aligned}
& + \frac{\epsilon}{\eta_n} \sum_{k=1}^n \sup_{\theta \in \Theta} |g_{k-1}(\theta) [\Delta_k^n Q]| \mathbf{1}_{\{\|\Delta Q^\epsilon\|_k^* \leq \delta\}} \\
& =: J_n^{(1)} + J_n^{(2)}.
\end{aligned}$$

As for  $J_n^{(1)}$ : it follows from the assumption A1 that, for  $Y^{n,\epsilon}$  given in Lemma 1,

$$\begin{aligned}
|J_n^{(1)}| &= \frac{1}{\eta_n} \sum_{k=1}^n \int_{t_{k-1}^n}^{t_k^n} \sup_{\theta \in \Theta} |g(Y_s^{n,\epsilon}) [b(X_s, \theta_0) - b(Y_s^{n,\epsilon}, \theta_0)]| \, ds \cdot \mathbf{1}_{\{\|\Delta Q^\epsilon\|_k^* \leq \delta\}} \\
&\lesssim \left(1 + \sup_{t \in [0,1]} |X_t|\right)^C \cdot \frac{1}{n\eta_n} \sum_{k=1}^n \sup_{t \in (t_{k-1}^n, t_k^n]} |X_t - Y_t^{n,\epsilon}| \mathbf{1}_{\{\|\Delta Q^\epsilon\|_k^* \leq \delta\}} \\
&\lesssim \left(1 + \sup_{t \in [0,1]} |X_t|\right)^C \cdot (\|X - X^0\|_* + \|X^0 - Y^{n,\epsilon}\|_*) \cdot \frac{1}{n\eta_n} \sum_{k=1}^n \mathbf{1}_{\{\|\Delta Q^\epsilon\|_k^* \leq \delta\}}
\end{aligned}$$

Since

$$\mathbb{E} \left| \frac{1}{n\eta_n} \sum_{k=1}^n \mathbf{1}_{\{\|\Delta Q^\epsilon\|_k^* \leq \delta\}} \right| = 1 \quad \Rightarrow \quad \frac{1}{n\eta_n} \sum_{k=1}^n \mathbf{1}_{\{\|\Delta Q^\epsilon\|_k^* \leq \delta\}} = O_p(1), \quad (5.31)$$

and Lemma 1 we have that

$$|J_n^{(1)}| \xrightarrow{\mathbb{P}} 0.$$

As for  $J_n^{(2)}$ : it follows for some  $C > 0$  that

$$\begin{aligned}
|J_n^{(2)}| &\lesssim \sup_{\theta \in \Theta, t \in [0,1]} (1 + |X_t|)^C \cdot \frac{n\epsilon}{\eta_n} \cdot \sum_{k=1}^n \frac{1}{n} \sup_{s \in (t_{k-1}^n, t_k^n]} |Q_s - Q_{t_{k-1}^n}| \mathbf{1}_{\{\|\Delta Q^\epsilon\|_k^* \leq \delta\}} \\
&\leq O_p \left( n\epsilon \sup_{s \in (0, \Delta_n]} |Q_s| \right),
\end{aligned}$$

since  $\sup_{s \in (t_{k-1}^n, t_k^n]} |Q_s - Q_{t_{k-1}^n}| \sim^d \sup_{s \in (0, \Delta_n]} |Q_s|$  and (5.31). Moreover, noticing under  $Q1[\gamma]$  that

$$\sup_{s \in (0, \Delta_n]} |Q_s| = o_p(\Delta_n^\gamma),$$

we obtain that

$$J_n^{(2)} = O_p(n\epsilon\Delta_n^\gamma) \rightarrow 0.$$

This completes the proof of (5.28).  $\square$

Now, putting

$$\bar{\Psi}_{n,\epsilon}(\theta) := \epsilon^2 \eta_n^{-1} \left( \tilde{\Psi}_{n,\epsilon,\delta}(\theta) - \tilde{\Psi}_{n,\epsilon,\delta}(\theta_0) \right); \quad F(\theta) := \int_0^1 |b(X_t^0, \theta) - b(X_t^0, \theta_0)|^2 \, dt,$$

and using Lemma 8, we can easily show that

$$\sup_{\theta \in \Theta} |\bar{\Psi}_{n,\epsilon}(\theta) - F(\theta)| \xrightarrow{\mathbb{P}} 0.$$

This and the identifiability condition A3 yield the consistency:  $\tilde{\theta}_{n,\epsilon,\delta} \xrightarrow{\mathbb{P}} \theta_0$ .

Suppose that the Lévy process  $Q$  is of the form

$$Q_t = at + cW_t + \int_0^t \int_{|z| \leq 1} z \tilde{N}(dt, dz) + \int_0^t \int_{|z| > 1} z N(dt, dz),$$

where  $a \in \mathbb{R}^d$ ,  $c \geq 0$ ,  $W$  is a  $d$ -dimensional Wiener process,  $N$  is a Poisson random measure associated with jumps of  $Q$ , and  $\tilde{N}(dt, dz) = N(dt, dz) - \nu(dz)dt$ , and note that, for any  $\delta > 0$ ,

$$Q_t = \alpha_{\delta/\epsilon} t + W_t + \int_0^t \int_{|z| \leq \delta/\epsilon} z \tilde{N}(dt, dz) + \int_0^t \int_{|z| > \delta/\epsilon} z N(dt, dz),$$

where  $\alpha_{\delta/\epsilon} = a - \int_{\delta/\epsilon < |z| \leq 1} z \nu(dz)$ .

Hereafter, we put

$$\zeta := \delta/\epsilon.$$

Using the same notation as in the previous theorem, the proof ends if we show that

$$(\sigma(\zeta)\epsilon)^{-1} G_{n,\epsilon,\delta}(\theta_0) \xrightarrow{\mathcal{D}} \int_0^1 \nabla_{\theta} b(X_t^0, \theta_0) [dB_t]. \quad (5.32)$$

Note that

$$\begin{aligned} (\sigma(\zeta)\epsilon)^{-1} G_{n,\epsilon,\delta}(\theta_0) &= (\sigma(\zeta)\epsilon)^{-1} \sum_{k=1}^n \nabla_{\theta} b_{k-1}(\theta_0) \left[ \int_{t_{k-1}^n}^{t_k^n} \{b(X_s, \theta_0) - b_{k-1}(\theta_0)\} dt \right] \mathbf{1}_{\{\|\Delta Q^{\epsilon}\|_k^* \leq \delta\}} \\ &\quad + \sigma^{-1}(\zeta) \sum_{k=1}^n \nabla_{\theta} b_{k-1}(\theta_0) [a_{\delta} \Delta_n + \Delta_k^n W] \mathbf{1}_{\{\|\Delta Q^{\epsilon}\|_k^* \leq \delta\}} \\ &\quad + \sigma^{-1}(\zeta) \sum_{k=1}^n \nabla_{\theta} b_{k-1}(\theta_0) \left[ \int_{t_{k-1}^n}^{t_k^n} \int_{|z| \leq \zeta} z \tilde{N}(dt, dz) \right] \\ &=: I_{n,\epsilon,\delta}^{(1)} + I_{n,\epsilon,\delta}^{(2)} + I_{n,\epsilon,\delta}^{(3)}. \end{aligned}$$

As for  $I_{n,\epsilon,\delta}^{(2)}$ : Since  $\mathbf{1}_{\{\|\Delta Q^{\epsilon}\|_k^* \leq \delta\}}$  is independent of  $\Delta_k^n W$  and  $X_{t_{k-1}^n}$ , we have that

$$\begin{aligned} \mathbb{E} \left| I_{n,\epsilon,\delta}^{(2)} \right| &\leq \sigma^{-1}(\zeta) \frac{1}{n} \sum_{k=1}^n \mathbb{E} |\nabla_{\theta} b_{k-1}(\theta_0) a_{\delta}| \mathbb{P}(\|\Delta Q\|_k^* \leq \zeta) \\ &\quad + \sigma^{-1}(\zeta) \sum_{k=1}^n \mathbb{E} |\nabla_{\theta} b_{k-1}(\theta_0) [\Delta_k^n W]| \mathbb{P}(\|\Delta Q\|_k^* \leq \delta) \end{aligned}$$

$$\begin{aligned}
&\leq \sigma^{-1}(\zeta) \sum_{k=1}^n \left( \mathbb{E} \int_{t_{k-1}^n}^{t_k^n} \text{trace}(\nabla_{\theta} b_{k-1}^{\otimes 2}(\theta_0)) \, dt \right)^{1/2} \mathbb{P}(\|\Delta Q\|_k^* \leq \zeta) \\
&= O\left(\sigma^{-1}(\zeta) e^{-\lambda(\zeta)\Delta_n}\right).
\end{aligned}$$

The last equality is due to (5.26). Therefore we have

$$\mathbb{E} |I_{n,\epsilon,\delta}^{(2)}| = O\left(\left(\sigma(\zeta) \sum_{k=0}^{\infty} \frac{(\Delta_n \lambda(\zeta))^k}{k!}\right)^{-1}\right). \quad (5.33)$$

For an integer  $M$  such that  $1/M \leq \rho$ , we see that,

$$\sigma(\zeta)(\Delta_n \lambda(\zeta))^M = \left(\sigma^{1/M}(\zeta) \Delta_n \lambda(\zeta)\right)^M \geq (\sigma^{\rho}(\zeta) \Delta_n \lambda(\zeta))^M \rightarrow \infty \quad (5.34)$$

due to the condition (4.4). Hence we have

$$\mathbb{E} |I_{n,\epsilon,\delta}^{(2)}| \rightarrow 0.$$

As for  $I_{n,\epsilon,\delta}^{(1)}$ : By the same argument as in the proof of Lemma 3.6,  $H_{n,\epsilon}^{(1)}(\theta_0)$  in [20], we can obtain the following inequality:

$$\begin{aligned}
|I_{n,\epsilon,\delta}^{(1)}| &\lesssim \frac{1}{n\epsilon\sigma(\zeta)} \frac{1}{n} \sum_{i=1}^n |\nabla_{\theta} b_{k-1}(\theta_0)| \cdot |b_{k-1}(\theta_0)| \\
&\quad + \frac{1}{n\sigma(\zeta)} \sum_{i=1}^n |\nabla_{\theta} b_{k-1}(\theta_0)| \sup_{s \in (t_{k-1}^n, t_k^n]} |Q_t - Q_{t_{k-1}^n}| \mathbf{1}_{\{\|\Delta Q^{\epsilon}\|_k^* \leq \delta\}}
\end{aligned}$$

Using the same estimates for  $J_n^{(2)}$  in the proof of Lemma 8, we have that

$$|I_{n,\epsilon,\delta}^{(1)}| = O_p\left(\frac{1}{n\epsilon\sigma(\zeta)}\right) + o_p\left(\sigma^{-1}(\zeta) e^{-\lambda(\zeta)\Delta_n}\right),$$

Then, by the same estimates as for (5.33) and (5.34) above, we see that  $|I_{n,\epsilon,\delta}^{(1)}| \xrightarrow{\mathbb{P}} 0$ .

As for  $I_{n,\epsilon,\delta}^{(3)}$ : Let

$$L_t^{\delta} := \sigma^{-1}(\zeta) \int_0^t \int_{|z| \leq \delta} z \tilde{N}(dt, dz).$$

Thanks to Theorem 2 by Asmussen and Rosinski [1], it follows under the assumption (4.6) that there exists a Wiener process  $B$ , independent of  $W$ , such that

$$L^{\zeta} \xrightarrow{\mathcal{D}} B \quad \text{in } \mathbb{D}[0, 1],$$

as  $\delta \rightarrow 0$ . Moreover, since  $\|Y^{n,\epsilon} - X^0\|^*$  a.s. by Lemma (1), we have a joint convergence

$$(L^{\zeta}, Y^{n,\epsilon}) \xrightarrow{\mathcal{D}} (B, X^0) \quad \text{in } \mathbb{D}[0, 1].$$



as  $\epsilon, \delta \rightarrow 0$  and  $n \rightarrow \infty$ . Hence it follows from Theorem 5.16 by Jacod and Shiryaev [10] that

$$\begin{aligned} I_{n,\epsilon,\delta}^{(3)} &= \sum_{k=1}^n \nabla_{\theta} b_{k-1}(\theta_0) \left[ \Delta_k^n L^{\zeta} \right] = \int_0^1 \nabla_{\theta} b(Y_t^{n,\epsilon}, \theta_0) dL_t^{\zeta} \\ &\xrightarrow{\mathcal{D}} \int_0^1 \nabla_{\theta} b(X_t^0, \theta_0) dB_t \end{aligned}$$

This completes the proof of (5.32), and the statement is proved.

## Acknowledgement

This research has been partially supported by the Ministry of Education, Science, Sports and Culture, Grant-in-Aid for Young Scientists (B), no.24740061, 2013–2014; Scientific Research (A), no. 25245033, 2014; and Japan Science and Technology Agency, CREST.

## References

- [1] Asmussen, S. and Rosinski, J. (2001). Approximations of small jumps of Lévy processes with a view towards simulation, *J. Appl. Prob.*, **38**, 482–493.
- [2] Bichteler, K. and Jacod, J. (1986) Calcul de Malliavin pour les diffusions avec sauts: existence d’une densité dans le cas unidimensionnel. *Lecture Notes in Math*, 132–157. Springer-Verlag, Berlin.
- [3] Boukai, B. (1990). An explicit expression for the distribution of the supremum of Brownian motion with a change point. *Comm. Statist. Theory Methods*, **19**, (1), 31–40.
- [4] Cont, R. and Tankov, P. (2004). *Financial modelling with jump processes*. Chapman & Hall/CRC, Boca Raton, FL.
- [5] Doob, J. L. (1949). Heuristic approach to the Kolmogorov-Sminorv Theorems. *Ann. Math. Statistics*, **20**, 393–403.
- [6] Genon-Catalot, V. (1990). Maximum contrast estimation for diffusion processes from discrete observations, *Statistics*, **21**. 99–116.
- [7] Genon-Catalot, V. and Jacod, J. (1993). On the estimation of the diffusion coefficient for multidimensional diffusion process, *Ann. Inst. Henri Poincaré Prob. Statist.* **29**, 119–151.
- [8] Gloter, A. and Sørensen, M. (2009). Estimation for stochastic differential equations with a small diffusion coefficient, *Stochastic Process. Appl.*, **119**, 679–699.

- [9] Ibragimov, I. A. and Has'minskii, R. Z. (1981) *Statistical Estimation*, Springer-Verlag, Berlin.
- [10] Jacod, J. and Shiryaev, A. N. (2003). *Limit theorems for stochastic processes*. Second edition. Springer-Verlag, Berlin.
- [11] Joulin, A. (2007). On maximal inequalities for stable stochastic integrals. *Potential Anal.*, **26**, (1), 57–78.
- [12] Kunitomo, N. and Takahashi, A. (2001). The asymptotic expansion approach to the valuation of interest rate contingent claims. *Math. Finance*, **11**, 117–151.
- [13] Kutoyants, Yu. A. (1984). *Parameter estimation for stochastic processes*. Heldermann, Berlin.
- [14] Kutoyants, Yu. A. (1994). *Identification of dynamical systems with small noise*. Kluwer, Dordrecht.
- [15] Laredo, C.F. (1990). A sufficient condition for asymptotic sufficiency of incomplete observations of a diffusion process, *Ann. Statist.*, **18**, 1158–1171.
- [16] Long, H. (2009) Least squares estimator for discretely observed Ornstein-Uhlenbeck processes with small Lévy noises, *Statist. Prob. Letters*, **79**, 2076–2085.
- [17] Long, H. (2010). Parameter estimation for a class of stochastic differential equations driven by small stable noises from discrete observations, *Acta Mathematica Scientia*, **30B**, 645–663.
- [18] Ma, C. (2010). A note on “Least squares estimator for discretely observed Ornstein-Uhlenbeck processes with small Lévy noises”, *Statist. Prob. Letters*, **80**, 1528–1531.
- [19] Masuda, H. (2013). Convergence of Gaussian quasi-likelihood random fields for ergodic Levy driven SDE observed at high frequency. *The Annals of Statistics*, **41**, 1593–1641.
- [20] Long, H; Shimizu, Y. and Sun, W. (2013). Least squares estimators for discretely observed stochastic processes driven by small Lévy noises. *J. Multivariate Analysis*, **116**, 422–439.
- [21] T. Ogihara, N. Yoshida, Quasi-likelihood analysis for the stochastic differential equation with jumps. *Stat. Inference Stoch. Process.*, **14**, 189–229.
- [22] Pavlyukevich, I. (2008). Ruin probabilities of small noise jump-diffusions with heavy tails. *Appl. Stoch. Models Bus. Ind.* **24**, (1), 65–82.
- [23] Pavlyukevich, I. (2011). First exit times of solutions of stochastic differential equations driven by multiplicative Lévy noise with heavy tails. *Stoch. Dyn.*, **11**, (2–3), 495–519.

- [24] Protter, P. E. (2004). *Stochastic integration and differential equations*. Second edition. Springer-Verlag, Berlin.
- [25] Pruitt, W. E. (1981). The growth of random walks and Lévy processes. *Ann. Probability*, **9**, (6), 948–956.
- [26] Shimizu, Y. (2006). M-estimation for discretely observed ergodic diffusion processes with infinitely many jumps. *Statist. Infer. Stochastic Proc.*, **9**, (2), 179–225.
- [27] Shimizu, Y. (2010). Threshold selection in jump-discriminant filter for discretely observed jump processes. *Statist. Methods and Appl.*, **19**, (3), 355–378.
- [28] Shimizu, Y. and Yoshida, N. (2006) Estimation of parameters for diffusion processes with jumps from discrete observations. *Statist. Infer. Stochastic Proc.*, **9**, (3), 227–277.
- [29] Sørensen, M. (2000). Small dispersion asymptotics for diffusion martingale estimating functions, Preprint No. 2000-2, Department of Statistics and Operation Research, University of Copenhagen, Copenhagen.
- [30] Sørensen, M. and Uchida, M. (2003). Small diffusion asymptotics for discretely sampled stochastic differential equations, *Bernoulli*, **9**, (6), 1051–1069.
- [31] Takahashi, A. (1999). An asymptotic expansion approach to pricing contingent claims. *Asia-Pacific Financial Markets*, **6**, 115–151.
- [32] Takahashi, A. and Yoshida, N. (2004). An asymptotic expansion scheme for optimal investment problems. *Stat. Inference Stoch. Process.*, **7**, 153–188.
- [33] Uchida, M. Estimation for discretely observed small diffusions based on approximate martingale estimating functions, *Scand. J. Statist.* **31** (2004) 553–566.
- [34] Uchida, M. (2008). Approximate martingale estimating functions for stochastic differential equations with small noises. *Stochastic Process. Appl.*, **118**, 1706–1721.
- [35] Uchida, M. and Yoshida, N. (2004). Asymptotic expansion for small diffusions applied to option pricing. *Stat. Inference Stoch. Process.*, **7**, 189–223.
- [36] Yoshida, N. (1992). Asymptotic expansion of maximum likelihood estimators for small diffusions via the theory of Malliavin-Watanabe. *Probab. Theory Relat. Fields*, **92**, 275–311.
- [37] Yoshida, N. Asymptotic expansion for statistics related to small diffusions. *J. Japan Statist. Soc.*, **22**, 139–159.
- [38] Yoshida, N. (2011). Polynomial type large deviation inequalities and quasi-likelihood analysis for stochastic differential equations. *Ann. Inst. Statist. Math.*, **63**, (3), 431–479